Towards understanding the evolution of *Banana bunchy top virus* and the detection of associated badnaviruses

A thesis submitted in partial fulfilment of the requirements for the Degree of

Doctor of Philosophy

at the

University of Canterbury, New Zealand

Daisy Stainton

2015

Table of contents

Acknowledgements

A huge thank you to all the people who have helped make this thesis journey possible.

Firstly, **Arvind Varsani**, who is truly an amazing supervisor. His passion and enthusiasm for science are contagious, and his continuous support, guidance and patience are invaluable.

My co-supervisors; **Dave Collings** and **Darren Martin**, who always found the time to lend their expertise.

To **John Thomas** for not only generously welcoming us into his Brisbane Lab when ours was closed, but also for access to a worldwide banana collection, from which the majority of the samples in this thesis were sourced.

Also a huge thank you to all the people who collected plant samples

To all of the following for the big and the little things, the coffee and cake breaks, laughs, skype calls and office drawings:

To my parents **Barbara** and **Ross**, and to the rest of my family **Jack**, **Jane**, **Billy**, **Aunty Jacq** and **Grandma**, thank you. Thank you for all your continued support and encouragement, from near and far, and for always believing in me.

To all my lab mates in the Varsani lab and the extras collected along the way. With special thanks to **Simona** (the tall to my short), **Ryan**, **Anisha**, **Matt**, **Dorien**, **Jacqui**, **Katherine, Laurel** and **Jei**.

Abstract

Bananas are an important subsistence and export crop with over 130 tonnes of banana produced annually. Domesticated banana are triploid and sterile and it is thought that wild diploid bananas were domesticated between 7,000 and 10,000 years ago somewhere in the region of Indonesia, Philippines, New Guinea or the Southeast Asia peninsula. Triploid bananas were subsequently moved around the world through vegetative propagules and more recently through tissue culture.

A number of diseases are associated with bananas, including banana bunchy top disease (BBTD) and banana streak disease (BSD), both caused by DNA viruses. BBTD is caused by *Banana bunchy top virus* (BBTV), a single-stranded DNA (ssDNA) virus (genus *Babuvirus*; family *Nanoviridae*). Characteristic symptoms of BBTD are severe stunting of the plant and bunching of the leaves, with a green dot-dash streak on the underside of the leaves. Both viruses can cause severe crop loss and are of concern to banana growing regions. BSD is caused by a number of banana-infecting badnaviruses species (*Badnavirus*; *Caulimoviridae*), all of which contain a double-stranded DNA (dsDNA) genome. Characteristic symptoms of BSD include chlorotic and necrotic streaks on the leaves, which are not always seen across all leaves, or even across an entire leaf.

BBTV is a multi-component virus which is transmitted by the banana aphid *Pentalonia nigronervosa*. It is the type member of the babuviruses, which contains two other accepted species, *Cardamom bushy dwarf virus* (CBDV) and *Abaca bunchy top virus* (ABTV), which infect large cardamom and *Musa* spp, respectively. It is generally accepted that the genome of BBTV consists of six different components, with each component individually encapsidated. BBTV is able to evolve through recombination, reassortment and mutation. The large numbers of sequenced BBTV components in Chapter Two and Three (927), and publically available component sequences, have allowed for in-depth analyses into the diversity and evolution of this important plant pathogen. Our analysis shows that both reassortment and recombination play a significant role in the evolution of BBTV. Additionally, we found high genetic diversity in two

geographic regions, the Southeast Asian / Far East region and the Indian subcontinent. Phylogenetic analysis of all full genomes, with recombinant regions and reassorted components removed, identified that the BBTV genomes circulating in the majority of countries have likely originated from single founder populations.

All nanoviruses contain two common regions which are common across species. The common region stem-loop (CR-SL) is involved in initiation of replication and is recognised by the replication-associated protein (Rep), encoded in the DNA-R component. The Rep recognises cognate components through iteron sequences in the CR-SL and nicks the component at the nonanucleotide motif to initiate replication. The common region major (CR-M) is involved in secondary strand replication. The large amount of sequence data generated in Chapters Two and Three, and a recent deposition of a large number of CBDV sequences, allowed a comprehensive analysis of the common regions of the babuviruses. All CR-SL across the three species showed high similarity including the iteron sequences. The CR-M regions however were not similar across the species. Alphasatellites, which only encode a Rep, have previously been found associated with some babuvirus, nanovirus and geminivirus isolates. Therefore these were also analysed for common regions. The babuvirus alphasatellites were more similar to the other alphasatellites than to the babuvirus components.

Both BBTV and banana-infecting badnaviruses are DNA viruses which infect the same host species, therefore the banana material that was collected for BBTV analyses was also screened for three species of banana-infecting badnaviruses. Species specific screening primers were designed for *Banana streak MY virus* (BSMYV), *Banana streak OL virus* (BSOLV) and *Banana streak GF virus* (BSGFV) across the movement protein motif located on open reading frame (ORF) three. With these primers we had a higher detection rate, for all three species, compared to previously published primers for banana-infecting badnaviruses. Eighty-two banana samples from 11 countries were found to be positive for at least one badnavirus, with the majority containing BSMYV. Of the 82 positive samples, 51 samples were also positive for BBTV. Banana-infecting badnaviruses are able to exist in two forms, a circular episomal infective form

which is transmitted by a number of mealybug species and an endogenous form which is integrated into the banana genome and does not cause symptoms. However, a number of endogenous forms have been identified which are able to reactivate from the *Musa* genome into the episomal infectious form. As both forms are potentially able to cause infection, detection of either form is important in the identification of clean planting stock.

This PhD thesis investigated banana infecting DNA viruses, *Banana bunchy top virus* and three banana-infecting badnaviruses. In-depth analyses of the global diversity, evolution and the dispersal of BBTV were undertaken. The common regions of the babuviruses and alphasatellites were characterised and screening primers for BSMYV, BSOLV and BSGFV were designed.

Abbreviation list

ABTV	*Abaca bunchy top virus*
AG	Asian Group
BBrMV	*Banana bract mosaic virus*
BBTD	Banana bunchy top disease
BBTV	*Banana bunchy top virus*
BCTIV	*Beet curly top Iran virus*
BIB	banana-infecting badnaviruses
BMD	banana mosaic disease
BMLCV	*Black medic leaf roll virus*
BSCAV	Banana streak CA virus
BSD	Banana streak disease
BSGFV	*Banana streak GF virus*
BSIMV	Banana streak IM virus
BSMYV	*Banana streak MY virus*
BSOLV	*Banana streak OL virus*
BSUAV	Banana streak UA virus
BSUIV	Banana streak UI virus
BSULV	Banana streak UL virus
BSUMV	Banana streak UM virus
BSVNV	*Banana streak VN virus*
BWAP	Banana wilt associated phytoplasma
CBDV	*Cardamom bushy dwarf virus*
cccDNA	circular covalently closed DNA
clink	cell-cycle link protein gene
Clink	cell-cycle link protein
CMV	*Cucumber mosaic virus*
cp	capsid protein gene
CP	capsid protein
CRESS	circular Rep-encoding ssDNA
CR-M	common region major
CR-SL	common region stem-loop
dsDNA	double-stranded DNA
ea	endogenous and activatable
eBIB	endogenous banana-infecting badnavirus
ECSV	*Eragrostis curvula streak virus*
ELISA	enzyme-linked immunosorbent assay
FBNSV	*Faba bean necrotic stunt virus*
FBNYV	*Faba bean necrotic yellows virus*
FBYLV	*Faba bean yellow leaf virus*
Foc	*Fusarium oxysporum* f.sp. *Cubense*
hDNA	heterogenous length high molecular weight double stranded DNA
HrCTV	*Horseradish curly top*

ICTV	International Committee for Taxonomy of Viruses
INPACT	in-plant activation
kb	kilobase
LAMP	Loop-mediated isothermal amplification
MDV	*Milk vetch dwarf virus*
ML	maximum-liklihood
mp	movement protein gene
MP	movement protein
MSV	*Maize streak virus*
NGS	next-generation sequencing
nm	nanometre
nsp	nuclear-shuttle protein gene
NSP	nuclear-shuttle protein
nt	nucleotide
ORF	open reading frame
PBGS	paper-based gene sensor
PCR	polymerase chain reaction
PNYDV	*Pea necrotic yellow dwarf virus*
PYSV	Pea yellow stunt virus
RCA	rolling circle amplification
RCR	rolling circle replication
RDR	recombination-dependent replication
RE	restriction enzyme
reg	regulatory gene
ren	repliction enhancer gene
rep	replication-associated protein gene
Rep	replication-associated protein
RNAi	RNA interference
RNAse H	ribonuclease H
RT	reverse transcriptase
SCSV	*Subterranean clover stunt virus*
SCTAV	*Spinach curly top Arizona virus*
sd	symptom determinant gene
SDT	Sequence demarcation tool
SPDs	specificity determinates
SPG	South Pacific Group
SpSCTV	*Spinach severe curly top*
ssDNA	single-stranded DNA
TAS	Triple antibody
TCTV	*Turnip curly top virus*
TPCTV	*Tomato pseudo-curly top virus*
trap/ss	silencing suppressor function gene
XcM	*Xanthomonas campestris* pv. *Musacearum*

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Deputy Vice-Chancellor's Office
Postgraduate Office

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 2

Stainton, D., Kraberger,S., Walters, M., Wiltshire,E.J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T.H., Aholelei,W., Taufa, L., Thomas, J.E, Collings, D.A., Martin, D.P., Varsani, A. (2012) Evidence of inter-component recombination, intra-component recombination and reassortment in Banana bunchy top virus. Journal of General Virology, 93, 1103-1119

Please detail the nature and extent (%) of contribution by the candidate:

Various people assisted with the field sampling of banana material.

All the molecular work was undertaken by Daisy Stainton. She worked with Arvind Varsani and Darren Martin to analyse the data. Darren Martin and Arvind Varsani double checked all the bioinformatics analysis.

Daisy Stainton wrote the manuscript and the rest of the authors provided comments / feedback to improve it.

Stainton contribution; 90%

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifys that:

- **The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work**
- **In cases where the candidate was the lead author of the co-authored work he or she wrote the text**

Name: Arvind Varsani Signature:  Date: *23 March 2015*

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 3

Stainton, D., Martin, D., Muhire, B., Lolohea, S., Halafihi, M., Lepoint, P., Blomme, G., Crew, K. S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D. A., Lemey, P., Harkins, G., Thomas, J. E., Varsani, A. (In review)
The global distribution of *Banana bunch top virus* reveals little evidence for recent, long distance dispersal events that have been human-mediated

Please detail the nature and extent (%) of contribution by the candidate:

Various people assisted with the field sampling of banana material.

All the molecular work was undertaken by Daisy Stainton. She worked with Arvind Varsani, Gordon Harkin, Philippe Lemey and Darren Martin to analyse the data. Darren Martin and Arvind Varsani double checked all the bioinformatics analysis.

Daisy Stainton wrote the manuscript and the rest of the authors provided comments / feedback to improve it.

Stainton contribution; 90%

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifys that:

- **The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work**
- **In cases where the candidate was the lead author of the co-authored work he or she wrote the text**

Name: Arvind Varsani Signature:  **Date:** ***23 March 2015***

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 4

Stainton, D., Martin, D.P., Thomas, J.E., Varsani, A. (In review) Defective molecules of Banana bunchy top virus and the characterisation of the common regions found in babuviruses and alphasatellite molecules. Archives of Virology

Please detail the nature and extent (%) of contribution by the candidate:

All the molecular work was undertaken by Daisy Stainton. She worked with Arvind Varsani and Darren Martin to analyse the data. Darren Martin and Arvind Varsani double checked all the bioinformatics analysis.

Daisy Stainton wrote the manuscript and the rest of the authors provided comments / feedback to improve it.

Stainton contribution; 90%

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifys that:

- **The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work**
- **In cases where the candidate was the lead author of the co-authored work he or she wrote the text**

Name: Arvind Varsani Signature:  **Date:** ***23 March 2015***

Chapter 1

Introduction

Contents

1.1 The history and importance of Bananas

Banana is both an important crop for subsistence farming and export, with over 100 million tonnes of banana and over 37 million tonnes of plantain grown worldwide (FAOSTAT, 2012). The banana plant (Order Zingiberales; Family Musaceae; Genus *Musa*) is a monocotyledon herbaceous plant with a pseudostem. Both banana and plantain are the fruit of the banana plants, with the sweet varieties generally referred to as banana or dessert banana and the cooking or starch bananas referred to as the plantains. Banana is also used in Africa to produce banana beer. As both plantain and bananas are banana fruit, the term banana will be used to encompass both from here on.

The Musaceae family consists of two genera, the *Ensete* and *Musa*, with bananas within the *Musa* genus. The *Musa* genus was previously split into five sections *Callimusa*, *Musa*, *Australimusa*, *Rhodochlamys* and *Ingentimusa*. These sections have recently been revised and have now been split into two sections, the *Musa* (incorporating the *Rhodochlamys*) and the *Callimusa* (incorporating the *Australimusa* and *Ingentimusa*) with 33 accepted species in the Musa and 37 in the *Callimusa* (Häkkinen, 2013). The vast majority of banana varieties consumed today are derived from the crossing of the wild bananas *Musa acuminata* and *M. balbisiana*, resulting in triploid sterile plants (33 chromosomes) which produce parthenocarpic (seedless) banana fruit. Another less common group of edible bananas are known as the Fe'i bananas which are thought to be cultivated from *Australimusa* (Jarret *et al.*, 1992) and are now part of the *Callimusa* section.

1.1.1 Genetics/nomenclature

Two wild seeded varieties of banana, *M. acuminata* and *M. balbisiana,* are the source of the majority of the cultivars grown today. Cultivars which also involve *Musa schizocarpa* and the *Australimusa* are also grown but are rare. Crossing of these diploid wild seeded varieties has resulted in triploid sterile cultivars with edible banana fruit. A nomenclature system was developed in 1955 which used a scoring system of visual trait differences between the wild cultivars *M. acuminata* (AA) and *M. balbisiana* (BB), such as flower and pseudostem colour, to classify the triploid bananas as AAA, AAB or ABB (Simmonds & Shepherd, 1955). This visual trait system is still widely used however genotyping is becoming more common. Four genome types have been identified, the most common are those varieties with A and B genomes (*M. acuminata* and *M. balbisiana* respectively), and the rarer S and T genomes (*M.*

schizocarpa and *Australimusa*). It is unknown whether the T genome is present in all *Callimusa* or only the species which were previously known as the *Australimusa*, therefore *Australimusa* is used for the T genome. Restriction fragment length polymorphism (RFLP) analysis has been used to compare specific cultivars (Carreel *et al.*, 2002; Jarret *et al.*, 1992; Nwakanma *et al.*, 2003). Flow cytometry and random amplified polymorphic DNA (RAPD) markers, specific to the A and B genomes, are used to identify the ploidy and genome composition of cultivars (Pillay *et al.*, 2000; Pillay *et al.*, 2006). Genomic *in situ* hybridisation (GISH), a modified fluorescence *in situ* hybridisation (FISH), where each chromosome is fluorescently marked as either A, B, S or T is a further method for identifying ploidy and genomes (D'Hont *et al.*, 2000). Interestingly a triploid cultivar (ABB), contained eight chromosomes from *M. acuminata*, rather than the expected 11, suggesting triploid bananas do not always contain a three way split of the 33 chromosomes (D'Hont *et al.*, 2000). To date only one banana genome has been fully sequenced, the *M. acuminata* genome (D'Hont *et al.*, 2012), and a *M. balbisiana* draft genome is available in public databases (Davey *et al.*, 2013).

1.1.2 Agriculture of bananas

The natural distribution of *M. balbisiana* is throughout Southeast Asia and *M. acuminata* is mostly found further south in the Malaysian/Indonesian/Philippines region (Figure 1.1) (Perrier *et al.*, 2011). Human mediated movement of these wild species occurred early in agricultural history with evidence of banana cultivation at a site in Papua New Guinea dated between 6950-6440 BP and a further early agricultural site in Cameroon, Africa dated 2750-2300 BP (Denham *et al.*, 2004; Mbida Mindzie *et al.*, 2001; Perrier *et al.*, 2011). These early agricultural sites suggest early cultivation of triploid cultivars without seeded fruit, with diploid to triploid cultivation events likely occurring multiple times throughout banana history, along with long distance movement of the cultivars (Perrier *et al.*, 2011). Bananas are now grown in over 130 countries (FAOSTAT, 2012), with estimates of the current number of banana cultivars ranging from 500 to over 1000. Local names of synonymous cultivars as well as similar local names for different cultivars cause difficulties in identifying the true numbers of cultivars. For example the Banana cultivar checklist which is an online repository contains over 5000 entries with both local name and cultivar synonyms where known (ProMusa, 2014).

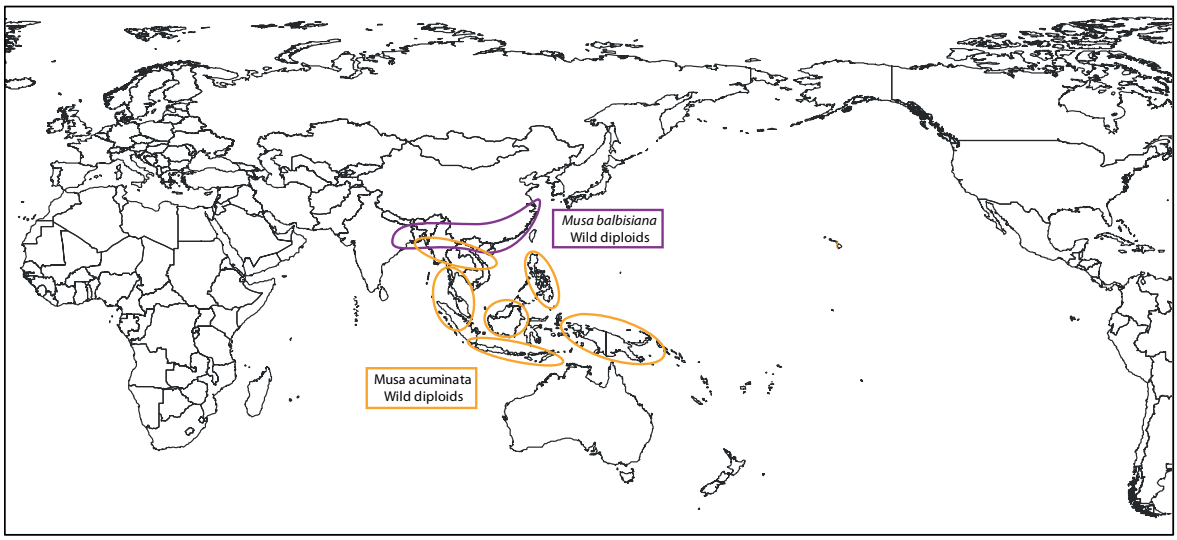


Figure 1.1: Natural geographic ranges of wild diploids of *M. acuminata* and *M. balbisiana* (Perrier *et al.*, 2011).

1.1.3 Agriculture of non-banana Musaceae

Abaca or Manila Hemp is produced from the leaves and pseuodostems of *Musa textilis* and is cultivated as a fibre crop rather than a fruit crop. The majority of Abaca fibre is produced in the Philippines, followed by Ecuador (FAOSTAT, 2012). Members of the *Ensete*, the other *Musaceae* genus, have seeded fruit like the wild Musa. Ethiopia is the only country which cultivates *Ensete* (*Ensete ventricosum*), where the pseudostem and corm are eaten as a staple or co-staple crop (Tsegaye & Struik, 2002).

1.2 Banana diseases

Bananas are important, both as a staple crop and as an economically important crop. Diseases which affect the yield of banana are caused by a number of pathogens including, fungi, bacteria, viruses and a phytoplasma. A selection of the main banana diseases are highlighted below.

1.2.1 Phytoplasma wilt disease

A banana wilt disease in Papua New Guinea causes wilting, leaf yellowing and leaf death, with necrotic streaks within the pseudostem. This disease is thought to be due to the phytoplasma, Banana wilt associated phytoplasma (BWAP) (Davis *et al.*, 2012). A coconut infecting phytoplasma is also present in Papua New Guinea, Cocos nucifera lethal yellowing phytoplasma 1, which has an identical 16S rRNA to BWAP suggesting that these are highly similar phytoplasma or that banana may be an alternate host for Cocos nucifera lethal yellowing phytoplasma (Davis *et al.*, 2012)

1.2.2 Bacterial wilt diseases

Banana wilt symptoms include wilted and yellow leaves, with uneven ripening of the banana fruit across the bunch. When the pseudostem is cut open, a characteristic yellow bacterial ooze exudes from the tissue and the tissue often displays necrotic brown streaks. Infected plants eventually die. The Banana wilt causing bacterial pathogens are spread by insects and mechanical inoculation, with the bacterium entering the plant through wounds. Two different bacterial species have been identified which cause bacterial wilt *Xanthomonas campestris* pv. *Musacearum* (XcM) and *Ralstonia solanacearum*.

The bacteria XcM has been found only in Africa and has become one of the most important constraints on *Musa* production in East Africa (Blomme *et al.*, 2014). Thought to be originally limited to Ethiopia, in 2001 it was identified in Uganda (Tushemereirwe *et al.*, 2004) and is now found across East and Central Africa (Mwangi & Nakato, 2009; Ndungo *et al.*, 2006; Reeder *et al.*, 2007). Africa has a recorded incidence which ranges from 12% to over 65% infection in crops, with XcM able to infect all cultivars (Mwangi & Nakato, 2009).

The bacterial wilt diseases Blood disease and Moko disease are caused by *R. solanacearum* (Peeters *et al.*, 2013). Currently Blood disease is only found in Indonesia, with the bacterial exudes white to reddish brown, rather than yellow (Kusumoto *et al.*, 2004). Moko disease, has a much larger geographical range, and has been found in the Philippines, Malaysia, Indonesia, Central and South America, and the Caribbean (PlantHealthAustralia, 2006; Zulperi & Sijam, 2014).

1.2.3 Fusarium wilt

Fusarium wilt is caused by the soil borne fungus, *Fusarium oxysporum* f.sp. *cubense* (Foc) which enters the plant through the roots causing symptoms similar to bacterial wilt, with yellowing of the leaves and wilting of the plant, splits in the pseudostem are also seen. Unlike the bacterial wilts, Fusarium wilt does not visibly infect the banana fruit and there is no pseudostem ooze. Fusarium wilt can result in rapid plant death and hence can have a devastating effect on banana plantations. For example Gros Michel was the main exported banana cultivar until the 1950s when Foc Race-1 of *F. oxysporum* had a devastating effect on the cultivar resulting in the switch to the resistant Cavendish variety. Cavendish however is susceptible to a further Foc Race, Foc tropical Race-4, which was first identified in Southeast Asia (Ploetz, 2006) and has now been identified in Africa and Jordan (García-Bastidas *et al.*, 2014) and is a concern as Cavendish is the major exported crop.

1.2.4 Black and yellow Sigatoka

The leaf fungal diseases, black Sigatoka (also known as black leaf streak) and yellow Sigatoka (also known as yellow leaf spot) are caused by the fungus *Mycosphaerella fijiensis* and *Mycosphaerella musicola*, respectively. Both Sigatoka diseases are found in all banana growing countries. The fungi cause black necrotic spots or streaks on the leaves, which are surrounded by yellow and can cover large areas of the leaf which reduces the photosynthetic ability of the plant. The diseases also result in uneven ripening of the fruit. Both fungi are

found in all banana growing regions of the world. Banana cultivars show large susceptibility differences, with *M. fijiensis* affected leaf surfaces ranging from 20-97% of the total leaf (Irish *et al.*, 2013). Differences in altitude preference of these fungal disease has been identified in Cameroon, with *M. musicola* favouring highlands and *M. fijiensis* favouring the lowlands (Mouliom-Pefoura *et al.*, 1996).

1.2.5 Banana bract mosaic disease

Banana bract mosaic disease is caused by *Banana bract mosaic virus* (BBrMV). BBrMV is a single-stranded positive sense RNA virus (Genus; *Potyvirus*, Family; *Potyviridae*), with rod shaped particles, which was first identified in the Philippines (Bateson & Dale, 1995; Thomas *et al.*, 1997). BBrMV is transmitted via infected planting material and vectored by several aphid species, including *Rhopalosiphum maidis* and *Aphis gossypii* (Magnaye & Espino, 1990). The disease is found in Asia and the South Pacific, and has recently been identified in Ecuador (Lava Kumar *et al.*, 2014; Quito-Avila *et al.*, 2013). Symptoms include dark brown/red mosaic patterns on the flowers, spindle shaped green/brown streaks on the leaves and once dead leaf sheathes are removed the exterior of the pseudostem shows strong streaking patterns (Magnaye & Espino, 1990). Yield losses of up to 70% have been recorded as reviewed in Lava Kumar *et al.* (2014).

1.2.6 Banana mosaic disease

Banana mosaic disease (BMD), also known as infectious chlorosis, causes a chlorotic mosaic leaf pattern and deformation of the banana fruit. The disease is caused by *Cucumber mosaic virus* (CMV) which infects a large number of plant species including Musa. CMV is a positive sense single-stranded RNA (Genus; *Cucumovirirus*, Family; *Bromoviridae*), with a genome that consists of three separate RNA strands which are encapsidated together. BMD is vectored by the aphid species *Aphis gossypii* and *A. craccivora*, and is also transmissible through mechanical inoculation (Dheepa & Paranjothi, 2010; Vishnoi *et al.*, 2013).

1.2.7 Banana streak disease

Banana streak disease, also known as chlorotic streak disease, is thought to be the most widely distributed banana virus in the world. The causative agents of banana streak disease are banana-infecting badnaviruses (BIB) (family; *Caulimoviridae)*, with four accepted species *Banana streak VN virus* (BSVNV), *Banana streak MY virus* (BSMYV); *Banana*

streak OL virus (BSOLV); *Banana streak GF virus* (BSGFV) and a further six putative species. These BIB cause chlorotic streaks on the leaves which turn necrotic however unusually not all leaves show symptoms and the symptoms can be sporadic across a single leaf, and pseudostem splitting is also seen (Lockhart, 1968). Symptoms can be mild or severe, with severe infections usually resulting in rapid plant death.

Badnaviruses are pararetroviruses and although they do not encode an integrase, they are able to integrate into the host genome. Badnaviruses have two forms, the episomal form where the circular double-stranded DNA (dsDNA) (~7.4 kilobase (kb)) is within a bacilliform shaped viral particle and causes disease, and the endogenous form where the viral DNA is integrated into the banana genome. The episomal banana badnaviruses are transmitted by a number of different mealybug species. Endogenous banana badnaviruses sequences have been identified which become episomal and infectious, these are thought to occur when the plant is under stress (Chabannes *et al.*, 2013; Côte *et al.*, 2010; Harper *et al.*, 1999; Iskra-Caruana *et al.*, 2010). BIB along with Banana bunchy top virus are both members of the top ten economically important viral diseases (Rybicki, 2014).

1.2.8 Banana bunchy top disease

Banana bunchy top disease (BBTD) is caused by a multi-component single-stranded DNA (ssDNA) virus, *Banana bunchy top virus* (BBTV) (Genus; *Babuvirus*, Family; *Nanoviridae*). Early reports suggest BBTD has been observed as early as the 1880s (Magee, 1927) and is now found in nearly all banana growing regions except the Americas. The virus is spread through infected suckers and the aphid species *Pentalonia nigronervosa*. BBTD causes stunting of the plant which is often severe, bunching of the leaves and a dot-dash green streak pattern on the underside of the leaf which extends from the midrib to the margin. Infected plants rarely produce bananas. BBTD is one of the most important viral diseases in the developing world as well as one of the most economically important diseases (Dale, 1987; Rybicki, 2014; Rybicki & Pietersen, 1999).

1.3 Single-stranded DNA viruses

The International Committee for Taxonomy of Viruses (ICTV) was formed to classify the ever expanding world of virus discovery, with the first virus report published in 1971 (Wildy (1971), as cited in King *et al.* (2011)). Currently viruses are classified into groups based on their genomes; double-stranded RNA, positive sense single-stranded RNA, negative sense

single-stranded RNA, dsDNA, and ssDNA. With each group further broken down, where possible, into; order, family, genus and species. The ssDNA viruses are classified into nine different families (Table 1.1); members of *Anelloviridae*, *Circoviridae* and *Parvoviridae* infect animals, *Parvoviridae* and *Bidnaviridae* infect insects; *Inoviridae* and *Microviridae* infect bacteria; *Geminiviridae* and *Nanoviridae* infect plants; *Spiraviridae* infect Archea. Inoviruses consist of viruses with a filamentous or rod-shaped morphology, spiraviruses have coil shaped particle, geminiviruses have twinned icosahedral particles, and all other families of ssDNA viruses are encapsidated in icosahedral virons. The members of the *Parvoviridae* consist of linear ssDNA genomes whereas the other families all have circular DNA genomes.

Table 1.1: Overview of the ICTV accepted single-stranded DNA families, including host organisms, morphology and genome information

Single-stranded DNA virus families	Host organisms	Capsid morphology	DNA Genome	Approximate genome length	Number of DNA segments in genome
Anelloviridae	Animals	Icosahedral	Circular	2-3.9 kb	1
Circoviridae	Animals	Icosahedral	Circular	1.7-2.3 kb	1
Parvoviridae	Animals and insects	Icosahedral	Linear	4-6.3 kb	1
Inoviridae	Bacteria	filamentous or rod-shaped	Circular	4.5-12.4 kb	1
Microviridae	Bacteria	Icosahedral	Circular	4.4-5.3 kb	1
Geminiviridae	Plants	twinned icosahedral	Circular	2.7 kb-5.4 kb	1-2
Nanoviridae	Plants	Icosahedral	Circular	6.4 kb-8 kb	6-8
Bidnaviridae	Insects	icosahedral	Linear	13 kb	4
Spiraviridae	Archaea	coil shaped	Circular	25 kb	1

1.3.1 Geminiviridae

Geminiviruses are responsible for a large number of economically important crop diseases and hence have been intensively studied. Hosts of geminiviruses include both monocotyledon and dicotyledon plants and they are transmitted by sap sucking insects; whitefly, leafhoppers and tree hoppers. Geminivirus particles are twinned icosahedral virions of approximately 18-30 nanometre (nm) which encapsidates a single circular ssDNA sequence of ~2.7 kb (Zhang *et al.*, 2001). The majority of the species are monopartite with a ~2.7 kb genome within a single twinned particle. Some species have bipartite genomes which consist of two twinned icosahedral particles each containing ~2.7 kb to make a total genome size of ~5.4 kb. Currently there are seven accepted genera in the family *Geminiviridae*; *Begomovirus*, *Mastrevirus*, *Curtovirus*, *Becurtovirus*, *Topocuvirus*, *Eragrovirus* and *Turncurtovirus* (King *et al.*, 2011; Varsani *et al.*, 2014b). Only members of the genus *Begomovirus* contain species which are bipartite, all other genera have members with monopartite genomes. The total number of genes encoded varies across the genera, with all members of the *Geminiviridae* encoding genes on the virion-sense and complementary-sense strands. As the geminiviruses have been much more extensively studied than the nanoviruses, in many instances the similarities between these two families have led to extrapolations of functions to nanoviruses.

1.3.1.1 Begomovirus

Begomoviruses contain over 250 accepted species, the largest number of species of all the geminivirus genera and infect a wide range of dicotyledonous plants with *Bean golden mosaic virus* representing the type member (Fauquet *et al.*, 2008). Typical symptoms include leaf curling, mosaic patterns, leaf crumpling and / or chlorosis. Begomoviruses can be either monopartite, with a single component, or bipartite with two components known as DNA-A and DNA-B. The monopartite genome is ~2.7 kb and the bipartite genome consists of two components each of which is ~2.7 kb of circular ssDNA. All begomoviruses are vectored by the whitefly *Bemisia tabaci*, however, due to substitutions in a region of the capsid protein gene (*cp*), *Abutilon mosaic virus* (AbMV) is no longer able to be insect transmitted and is only able to be vegetatively propagated (Fischer *et al.*, 2015; Höhnle *et al.*, 2001). The bipartite begomoviruses encode six to eight genes, whereas the monopartite encodes six. Both encode a *cp* gene, replication-associated protein gene (*rep*), movement protein gene (*mp*), a symptom determinant gene (*sd*), a transcription activator protein gene / silencing suppressor

function gene (*trap/ss*) and a replication enhancer gene (*ren*). Only the Old World bipartite DNA-A components contain a pre-coat gene, which is also present on the monopartite genome, both on the virion-sense strand. The DNA-B of all bipartite species contains two genes, the *mp* on the complementary-sense, and a nuclear-shuttle protein gene (*nsp*) on the virion-sense strand.

1.3.1.2 Mastrevirus

The *Mastrevirus* genus contains 29 ICTV accepted species, with *Maize streak virus* (MSV) as the type species. Mastreviruses infect either monocotyledonous or dicotyledonous plants and are all vectored by *Circadellidae* spp (leafhopper). Symptoms can include plant stunting, leaf yellowing, leaf reddening, striations and / or mosaic patterns, with symptoms often varying based on host. Mastreviruses have only been found in the Old World with the exception of mastrevirus sequences which have been found in sweet potatoes in Peru (Kreuze *et al.*, 2009) and a mastrevirus which has been isolated from dragonflies in Puerto Rico (Rosario *et al.*, 2013). Mastreviruses have four open reading frames (ORFs) which encode the *mp* and *cp* on the virion-sense, and two rep protein genes, *rep* and *repA* on the complementary-sense strand, with *repA* expressed from a single ORF (C1) whereas *rep* is expressed from a spliced ORF (C1:C2) (Dekker *et al.*, 1991; Mullineaux *et al.*, 1990; Wright *et al.*, 1997).

1.3.1.3 Curtovirus

The classification of the *Curtovirus* was recently revised, and consists of three species, the type species *Beet curly top virus*, *Spinach severe curly top* (SpSCTV) and *Horseradish curly top* (HrCTV) (Varsani *et al.*, 2014a). Symptoms include stunting and necrosis, vein swelling, leaf curling and yellowing and leaf distortion, with these species infecting a wide range of dicotyledonous plants. The vector of the *Curtovirus* species is the leafhopper *Circulifer tennellus* (Chen & Gilbertson, 2009). Curtoviruses encode up to seven genes, four on the complementary-sense strand *rep*, *ren*, *trap/ss* and *sd*, and three on the virion-sense *mp*, regulatory gene (*reg*) and *cp*.

1.3.1.4 Becurtovirus

There are currently only two species in the *Becurtovirus* genus, the *Beet curly top Iran virus* (BCTIV) and *Spinach curly top Arizona virus* (SCTAV) (Hernández-Zepeda *et al.*, 2013; Yazdi *et al.*, 2008). Symptoms include stunting, leaf rolling, yellowing, vein swelling and

deformation (Gharouni Kardani *et al.*, 2013; Hernández-Zepeda *et al.*, 2013). BCTIV infects dicot plants and is vectored by the leaf hopper *C. haemoceps*, currently the virus has only been identified in Iran (Gharouni Kardani *et al.*, 2013; Heydarnejad *et al.*, 2013; Yazdi *et al.*, 2008). SCTAV infects spinach and has only been identified in the United States of America, no vector has yet been identified (Hernández-Zepeda *et al.*, 2013). Becurtoviruses encode five genes, three on the virion-sense strand, *mp*, *cp*, and a potential *reg*, and two on the complementary-sense strand *rep*, and a *repA*.

1.3.1.5 Topocuvirus

The genus *Topocuvirus* contains a single species, *Tomato pseudo-curly top virus* (TPCTV) which infects dicotyledonous plants with the major hosts *Solanaceae* spp (Briddon *et al.*, 1996). Symptoms of TPCTV include leaf curling and are similar to those of *Beet curly top virus* (genus *Curtovirus*). TPCTV is the only species of Geminivirus which is vectored by a treehopper, with the species *Micrutalis malleifera* shown to transmit the virus (Simons & Coe, 1958). TPCTV contains six genes, four on the complementary-sense strand, *rep*, *ren*, possible *trap/ss* and an unknown gene, and two on the virion-sense strand, *mp* and *cp*.

1.3.1.6 Eragrovirus

Another genus with only one species is the *Eragrovirus*, with the sole member *Eragrostis curvula streak virus* (ECSV) (Varsani *et al.*, 2014b). ECSV has only been found in the grass species *Eragrostis curvula* and causes mild leaf streaks. The viral vector of this genus is currently unknown. ECSV contains two genes on each strand with the virion-sense strand containing a possible *mp* and *cp*, and the complementary-sense strand containing *rep* and possible *trap/ss*.

1.3.1.7 Turncurtovirus

The *Turncurtovirus* genus also only contains a single species, *Turnip curly top virus* (TCTV) which has only been identified in Iran (Briddon *et al.*, 2010; Razavinejad & Heydarnejad, 2013; Razavinejad *et al.*, 2013). The virus is vectored by the leafhopper species *Circulifer haematoceps* across a broad range of dicotyledonous host species, with symptoms including leaf swelling and cupping (Razavinejad & Heydarnejad, 2013). The TCTV genome contains six genes, four on the complementary-sense *ren*, *rep*, and a possible *trap/ss* and possible *sd*.

1.3.1.8 Satellite DNA molecules associated with geminiviruses

Monopartite begomoviruses are often associated with satellite molecules, alphasatellites and betasatellites, which are approximately half the size (~1300 nt) of the associated virus and are packaged separately. A small number of bipartite begomoviruses have also been identified to be associated with alphasatellite molecules (Jeske *et al.*, 2014; Paprotka *et al.*, 2010b; Romay *et al.*, 2010). A recent report has identified an alphasatellite molecule associated with a mastrevirus (Kumar *et al.*, 2014). Alphasatellites (formally DNA1) and betasatellites (formerly DNAβ) contain a single ORF and a stem-loop region involved in rolling circle replication. Alphasatellites have a *rep* and are able to replicate themselves, whereas betasatellites encode a gene for pathogenicity and therefore rely on the associated virus for replication. As only one gene is encoded, both alphasatillites and betasatellites require the associated virus for encapsidation. Betasatellites contain a common region adjacent to the stem-loop which is conserved across all betasatellites (Zhou, 2013), no common region has been identified in the alphasatellites.

Some species of begomoviruses are always found with betasatellites and without their presence no disease symptoms are seen within the host (Cui *et al.*, 2004; Guo *et al.*, 2008). Whereas in other species of begomoviruses, only some isolates have been identified with a betasatellite, with the presence of the betasatellite resulting in stronger disease symptoms (Li *et al.*, 2005). The majority of infections involving an alphasatellite show no changes in plant symptoms (Zhou, 2013), however, a small number have resulted in attenuation of symptoms (Idris *et al.*, 2011; Nawaz-ul-Rehman *et al.*, 2010). Alphasatellite molecules have been found associated with a number of nanoviruses, but to date betasatellites have only been found with the begomoviruses, however recently a single betasatellite isolate was identified associated with a mastrevirus (Kumar *et al.*, 2014).

1.3.2 Nanoviridae

There are two genera in *Nanoviridae* family; *Nanovirus* and *Babuvirus,* both vectored by aphids and which infect a range of plant hosts. All members have multi-component genomes, with the genome of each species split across a number of circular ssDNA strands, each approximately 1 kb (Figure 1.2, Table 1.2). Members of the *Nanovirus* genera infect dicotyledonous plants (legumes) whereas those of *Babuvirus* genera infect monocotyledonous plants (Table 1.2). Members of *Nanovirus* and *Babuvirus* also differ in the

total number of components which are generally accepted to constitute the genome, eight and six respectively, and the presence/absence of DNA-U1, DNA-U2, DNA-U3 and DNA-U4. With babuviruses containing DNA-U3 and nanoviruses containing DNA-U1, DNA-U2 and DNA-U4 (King *et al.*, 2011). The size of each component of the two genera also differs, with nanoviruses consisting of smaller components than the babuviruses (Table 1.2).

Within each genus overlaps in host species have been identified, for example *Abaca bunchy top virus* (ABTV) and BBTV both infect *Musa* spp, and *Pea necrotic yellow dwarf virus* (PNYDV) and *Milk vetch dwarf virus* (MDV) both infect *Pisum sativum* (Table 1.2). Therefore although natural host range, serological reactivity and vector species can be taken into account in species demarcation, the overall genome nucleotide (nt) sequence identity of <75% similarity can be used alone, for example Pea yellow stunt virus (PYSV) and Black medic leaf roll virus (BMLCV) have recently been proposed as new species based solely on the sequence threshold (Grigoras *et al.*, 2014; King *et al.*, 2011).

Table 1.2: Overview of all babuvirus and nanovirus species, including the size (nt) and presence/absence of components.

				Components										
				DNA-R	DNA-S	DNA-M	DNA-C	DNA-N	DNA-U1	DNA-U2	DNA-U3	DNA-U4	DNA-Uf1	DNA-Uf2
Species	Acronym	Host plant species	Country of sample collection	*rep*	*cp*	*mp*	*clink*	*nsp*	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
Nanovirus genus														
Subterranean clover stunt virus	SCSV	*Trifolium subterraneum*	Australia	1005	998	1001	991	1002	988	-	-	-	-	-
Faba bean necrotic yellows virus	FBNYV	*Vicia faba*	Spain Azerbaijan Ethiopia Iran Syria Egypt Morocco	1003	1006	992	999	986	996	1020	-	991	-	-
Faba bean necrotic stunt virus	FBNSV	*Phaseolus vulgaris* *Lens culinaris* *Vicia sativa*	Ethiopia Morocco Azerbaijan	1003	992	980	994	981	986	984	-	987	-	-
Pea necrotic yellow dwarf virus	PNYDV	*Pisum sativum*	Austria Germany	1002	981	988	988	993	978	985	-	979	-	-
Milk vetch dwarf virus	MDV	*Astragalus sinicus* *Vicia faba* *Pisum sativum* *Glycine mas*	China Japan	1001	997	985	990	977	989	981	-	991	-	-
Faba bean yellow leaf virus	FBYLV	*Vicia faba*	Ethiopia	1002	1001	980	995	1000	990	995	-	972	-	-
Unclassified Nanovirus														
Pea yellow stunt virus	PYSV	*Pisum sativum*	Austria Azerbaijan Sweden	1002	976	975	971	977	970	971	-	983	-	-
Black medic leaf roll virus	BMLRV	*Medicago lupulina* *Pisum sativum*	Austria Azerbaijan Sweden	1008	1017	1010	1017	1021	1011	1001	-	1006	-	-
Babuvirus genus														
Banana bunchy top virus	BBTV	*Musa* spp	Global banana regions (except the Americas)	1110	1076	1047	1017	1090	-	-	1064	-	-	-
Abaca bunchy top virus	ABTV	*Musa* spp	Phillipines, Malaysia	1099	1078	1076	1013	1073	-	-	1059	-	-	-
Cardamom bushy dwarf virus	CBDV	*Amomum subulatum*	India	1102	1086	1083	1027	1116	-	-	1088	-	1080	1078

Figure 1.2: A cartoon illustration of the components of the three babuviruses, BBTV, ABTV and CBDV, and a representative of the nanoviruses, FBNYV. The ORFs and both common regions are shown, with the common region major also known as the common region II in the *Nanovirus* genus.

1.3.2.1 Components

It is generally accepted that six components are integral in babuvirus infection and eight for nanovirus infection (Figure 1.2). These components range in size from, 971 nt – 1116 nt, with nanoviruses containing slightly fewer nucleotides in all components types which are shared with those of the babuviruses (Table 1.2). The nanovirus and babuvirus species share five component types, DNA-R, DNA-S, DNA-M, DNA-C and DNA-N, all which encode a protein of known function.

DNA-R encodes the replication-associated protein (Rep) (Burns *et al.*, 1995; Hafner *et al.*, 1997b; Harding *et al.*, 1993) which is involved in replicating all integral components and DNA-S encodes the capsid protein (CP) which encapsidates each circular ssDNA molecule in an icosahedral virion of 17-20 nm size (Wanitchakorn *et al.*, 1997). DNA-C encodes the cell-cycle link protein (Clink) which is involved in switching the plant host into S-Phase in order to increase replication of the other components (Aronson *et al.*, 2000; Hipp *et al.*, 2014; Lageix *et al.*, 2007; Wanitchakorn *et al.*, 2000a). DNA-M encodes the movement protein (MP) (Wanitchakorn *et al.*, 2000a) and DNA-N encodes the nuclear-shuttle protein (NSP) (Wanitchakorn *et al.*, 2000a). In BBTV, but not in CBDV or ABTV, a second internal ORF has been identified in DNA-R which transcribes a small mRNA with an unknown function (Beetham *et al.*, 1997).

Along with these five components, the babuviruses also contain DNA-U3, and the nanoviruses contain DNA-U1, DNA-U2, and DNA-U4, all of which have an unknown function. Two further components DNA-Uf1 and DNA-Uf2, have been identified with the babuvirus *Cardamom bushy dwarf virus* (CBDV), which may also constitute integral components (Mandal *et al.*, 2013). Within all the components there are two common regions, the common region stem-loop (CR-SL) and a second common region, the common region major (CR-M) (babuviruses) or the common region II (nanoviruses) (Figure 1.2). These regions are common across the integral components which constitute a genome.

1.3.2.2 Alphasatellites

Members of the *Nanoviridae* family have been found associated with satellite molecules known as alphasatellites. These additional molecules contain a *rep*-like gene and have been identified with the integral components of nanoviruses (Boevink *et al.*, 1995; Grigoras *et al.*, 2014; Katul *et al.*, 1998; Sano *et al.*, 1998) and the babuviruses, BBTV (Bell *et al.*, 2002; Fu

et al., 2009; Horser *et al.*, 2001b; Wu *et al.*, 1994; Yu *et al.*, 2012) and CBDVs (Mandal *et al.*, 2013). No alphasatellites have yet been reported to be associated with ABTV. A maximum-likelihood phylogenetic tree inferred using WAG+G+I amino acid model of substitution of the Rep sequences of the babuvirus and nanovirus alphasatellite and a subset of integral Reps encoded by the DNA-R components of both nanoviruses and babuviruses reveals that those of the alphasatellites are distinct from the integral Reps (Figure 1.3). Interestingly the integral Reps are more similar to the Reps of other integral species and genera than to the alphasatellite Reps that have been associated with them. The BBTV alphasatellite Reps are found in two separate clades with CBDV, the only other babuvirus alphasatellite, within the BBTV clade. The nanovirus alphasatellites found associated with the same integral hosts, rarely group together with the MDV alphasatellites present in nearly all nanovirus alphasatellite clades.

Although these alphasatellite molecules encode Reps, they are unable to replicate the integral components they are found with and are only able to replicate themselves (Horser *et al.*, 2001a; Timchenko *et al.*, 1999; Timchenko *et al.*, 2000). Alphasatellites are also found associated with begomoviruses (Jeske *et al.*, 2014; Paprotka *et al.*, 2010b; Romay *et al.*, 2010; Zhou, 2013) and recently also with a mastrevirus (Kumar *et al.*, 2014). To date only BBTV isolates which fall in the Asian Phylogenetic group have been identified with alphasatellites. There have been no reports of changes in virulence or disease symptoms with BBTV isolates which are associated with a BBTV alphasatellite.

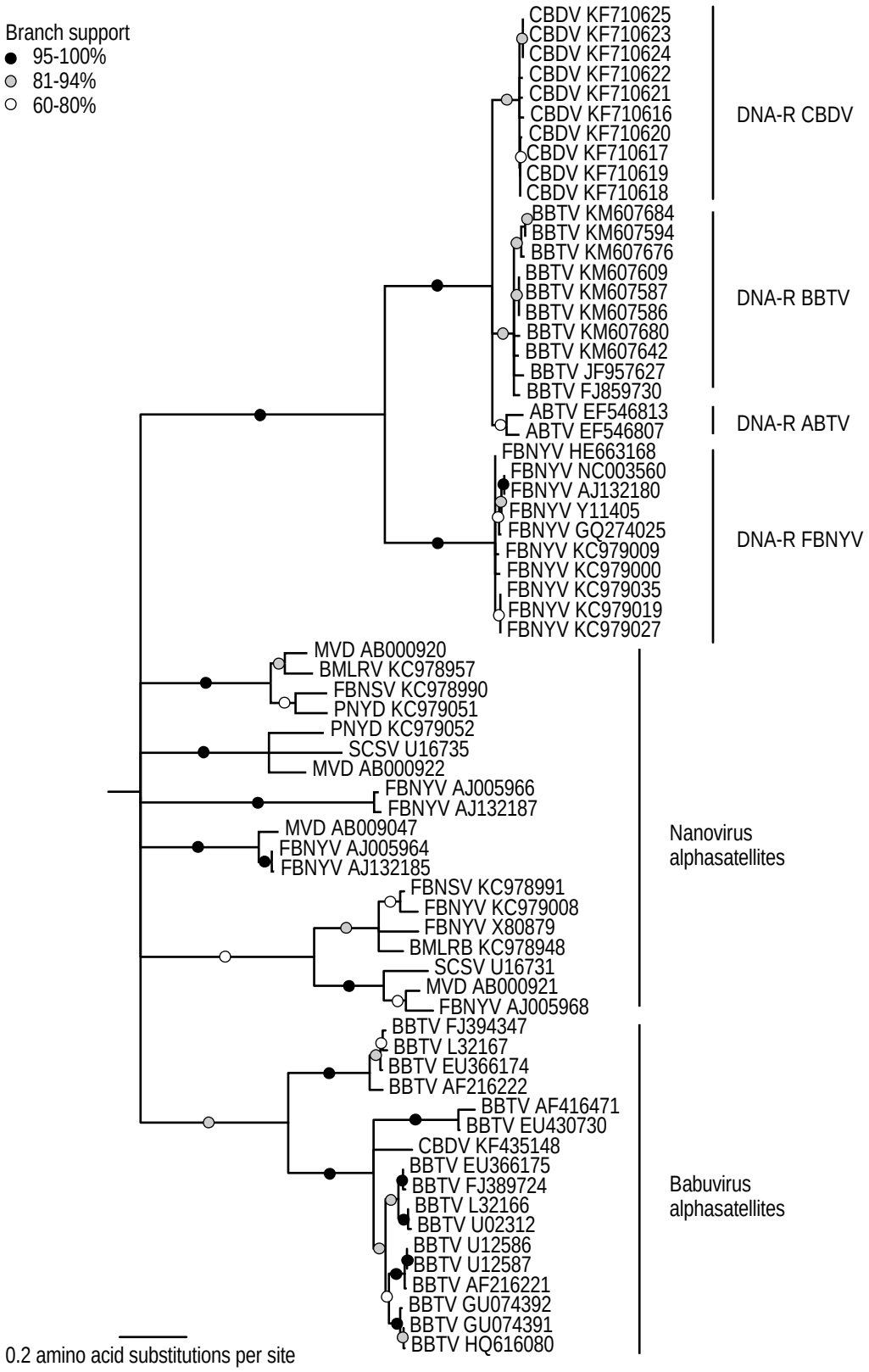


Figure 1.3: Maximum-likelihood amino acid phylogenetic tree (inferred using WAG G+I amino acid substitution model) of the Reps of all nanovirus associated satellites aligned with a subset of DNA-R Reps of babuvirus BBTV, CBDV and ABTV and a representative of the nanoviruses, FBNYV. The tree was rooted with geminivirus alphasatellites, and branches with bootstrap support of <60% have been collapsed in Mesquite v2.75 (http://mesquiteproject.org/).

1.3.2.3 Nanoviruses

There are six ICTV accepted members of the *Nanovirus* genus: *Subterranean clover stunt virus* (SCSV) (Boevink *et al.*, 1995; Chu & Helms, 1988) (type member), MDV (Sano *et al.*, 1998), *Faba bean necrotic yellows virus* (FBNYV) (Katul *et al.*, 1998), *Faba bean necrotic stunt virus* (FBNSV) (Grigoras *et al.*, 2009), PNYDV (Grigoras *et al.*, 2010a) and *Faba bean yellow leaf virus* (Abraham *et al.*, 2012). Two recently described putative nanoviruses, PYSV and BMLRV (Grigoras *et al.*, 2014), have also been identified. Although SCSV is the nanovirus type member, DNA-U2 and U4 have not been identified for this virus (Table 1.2)(Boevink *et al.*, 1995). Nanoviruses have currently been identified in Africa, Australia, Europe and Asia (Table 1.2). Some species such as SCSV and FBYLV have only been identified in a single country, Australia and Ethiopia respectively, whereas FBNYV has been found in Spain, Azerbaijan, Ethiopia, Iran, Syria, Egypt and Morocco (Table 1.2). Symptoms of nanoviruses include stunting, necrosis, leaf yellowing or reddening and leaf curling, all members of the nanovirus genus infect legumes.

1.3.2.4 Babuviruses

Babuviruses have three ICTV accepted members, BBTV (type member) (Harding *et al.*, 1991; Thomas & Dietzgen, 1991), ABTV (Sharman *et al.*, 2008) and CBDV (Mandal *et al.*, 2013). A further virus isolate, coconut foliar decay virus has had only a single component identified to date from Vanuatu and is vectored by a planthopper (*Myndus taffini*), rather than an aphid like the other *Nanoviridae* members (Randles *et al.*, 1986; Rohde *et al.*, 1990). Coconut foliar decay virus is currently classified as both; an unassigned member of the Nanoviridae and an assigned alphasatellite(King *et al.*, 2011), however, King *et al.* (2011) also states that it appears to be an alphasatellite sequence rather than an integral component and it is therefore likely to be reclassified. All three accepted species of babuviruses infect agriculturally important monocotyledonous plants, with CBDV infecting giant cardamom in the *Amomum* family, and ABTV and BBTV infecting *Musaceae*. Both BBTV and ABTV are able to infect banana and abaca plants (Magee, 1927; Sharman *et al.*, 2008).

1.3.2.4.1 *Cardamom bushy dwarf virus*

A single component of CBDV, DNA-R, was initially described by Mandal *et al.* (2004) as one of the components of the causal agent of the disease Foorkey which infects large Cardamom (*Amomum subulatum* Roxb.). Foorkey disease causes the cardamom plant to

become dwarfed and bushy with a reduction in yield and eventually causes the death of the infected plant (Mandal *et al.*, 2004). Isometric particles of nanovirus size were identified and the Rep gene was shown to be most similar to that of BBTV (Mandal *et al.*, 2004). However, it was not until 2013 that the other five components, and two potentially extra integral components (DNA-Uf1 and DNA-Uf2) were identified (Mandal *et al.*, 2013). Two species of aphid have been identified as vectors of CBDV, *Pentalonia nigronervosa* (Varma & Capoor, 1964) and *Mycromyzus kalimpongensis* basu (Basu & Ganguly, 1968; Mandal *et al.*, 2004), the latter was later renamed as *P. kalimpongensis* (Blackman & Eastop, 2008). Further *Pentalonia* spp. have been identified feeding on cardamom, however, whether these are able to transmit CBDV is currently unknown (Savory & Ramakrishnan, 2014b). Following the identification of the full genome of CBDV in 2013 a large scale study in India resulted in the sequencing of 163 isolates of CBDV (Savory & Ramakrishnan, 2014a). For each isolate the six components were sequenced using component specific primers, however DNA-Uf1 or DNA-Uf2 primers were not used therefore it is unknown if these components were also present (Savory & Ramakrishnan, 2014a). The data from this large scale study has been used to identify recombination and reassortment events as well as potential geographical reassortment hotspots in India (Savory & Ramakrishnan, 2014a; Savory *et al.*, 2014). Excluding the two additional components of CBDV, the integral components of ABTV and CBDV are similar to those of BBTV (Figure 1.2).

1.3.2.4.2 *Abaca bunchy top virus*

The symptoms of ABTV are similar to those of BBTV, with a bunched stunted appearance of the plant, leaf streaks and chlorotic margins (Magee, 1927; Sharman *et al.*, 2008). ABTV, like BBTV, is also spread by the banana aphid, *P. nigronervosa*. Unlike the other members of the babuviruses very little sequence data is available with only two complete genomes of ABTV sequenced, one from the Philippines and one from Malaysia (Sharman *et al.*, 2008).

1.3.2.4.3 *Banana bunchy top virus*

BBTD causes the leaves of the banana to bunch at the top and the leaves are more upright, giving the banana plant a distinctive bunchy top. The infected plant is often also severely stunted, especially if infected at a young age. Closer inspection of the underside of the leaves show a dark green dot-dash or morse code streaking pattern from the mid rib to the leaf edge with a 'hook' at the mid rib (Figure 1.4 A and B). The leaf margin may also be yellow.

Infected plants rarely produce bananas and if present the fruit is stunted and no longer marketable. BBTD has been reported from all growing regions of the world, except for the Americas, with outbreaks resulting in large scale crop loss. In Australia a BBTD outbreak in the 1920s nearly resulted in the loss of the entire banana industry with one banana growing district being reduced from 100 plantations to four (Pekin, 2006). The estimated cost of a large BBTD outbreak in Australia, with a 20% yield loss, is over 47.5 million in lost sales (AUD) (Pekin, 2006). Outbreaks of the disease not only have an economic impact on exports but also affect millions of people who rely on bananas as a staple crop.

In Australia BBTD was first identified around 1913 and was also known to be present in Sri Lanka, Egypt (1901), the Philippines and had been present in Fiji for at least 40 years (Fahmy, 1927; Magee, 1927). Early investigations were carried out in Australia to identify the causal agent which included nematodes, fungus and viruses as potential causes (Magee, 1927). Magee (1927) identified that BBTD was caused by a virus which was transmitted by the banana aphid *P. nigronervosa*, and was also able to be moved with infected suckers. After inoculation visual symptom identification from time of infection can range from 25 days to 85 days (Hooks *et al.*, 2008). Strict phyto-sanction recommendations were implemented in Australia following the outbreak which enabled the banana industry to control the disease in Australia (Dale, 1987; Magee, 1927). Total eradication of the disease in Australia has not been achieved despite the intensive control measures still enforced. To date no banana cultivars have been identified which are naturally resistant to BBTV, with destroying infected plants and reducing aphid numbers the main mechanisms of reducing the spread. In Australia infected plants are being injected with a cocktail of a fast acting insecticide and a slow acting herbicide in order to kill the aphids before they desert the dying plant (Parmenter & Thomas, 2014). Identification and disease management in developing counties is often at the local or farmer level, with access to clean planting material increasing the difficulties in controlling the disease.

Although the disease had been known about since the 1890s and was accepted to be the result of an aphid vectored virus, the causal agent had not been identified till the 1990s. Based on the yellowing symptoms, and that it was spread by an aphid it was initially thought to be a *leuteovirus*, an RNA virus (Matthews, 1981). It was not until the 1990s that a diagnostic test, enzyme-linked immunosorbent assay (ELISA), was available (Wu & Su, 1990) and that the causal agent was determined to be a ssDNA virus, with isometric particles of 18 to 20 nm in

size which contained approximately 1 kb of DNA (Harding *et al.*, 1991; Thomas & Dietzgen, 1991). Similarities between the genome size and particle morphology suggested BBTV was closer to the nanovirus SCSV and coconut foliar decay virus than to the geminiviruses (Harding *et al.*, 1991; Thomas & Dietzgen, 1991). BBTV is now found in all banana growing regions of the world, except the Americas with sequence data available from 22 countries (Figure 1.4C;Table 1.3).

The first component of BBTV to be sequenced was DNA-R (then known as DNA-1), which contained a circular ssDNA molecule which encodes a Rep, but did not contain an ORF of the size predicted for the CP of BBTV, suggesting BBTV contained a multi-component genome like SCSV (Harding *et al.*, 1993). A randomly primed library was used to clone and sequence sections of a potential second component, from these sections specific back-to-back primers amplified the full DNA-2 component (DNA-U3) (Burns *et al.*, 1995; Burns *et al.*, 1994). Comparison of DNA-1 and DNA-2 sequences identified a common region (the CR-M) in which back-to-back primers were designed to sequence a further four components, DNA-3 (DNA-S), DNA-4 (DNA-M), DNA-5 (DNA-C), DNA-6 (DNA-N) (Burns *et al.*, 1995; Burns *et al.*, 1994). Although all components contained an ORF only DNA-S, DNA-M, DNA-C and DNA-N contain an ORF in the virion sense along with a potential TATA box (CTATa/ta/tAt/aA) 5' of the ORF and a polyadenylation signal 3' of the ORF, with the exact distance from the ORFs variable across components (Burns *et al.*, 1995). The intergenic region containing the TATA box and also a G-box and an I-box, of DNA-N has been identified as a region which is required for strong promoter activity, with deletion resulting in no promoter activity (Dugdale *et al.*, 1998). Although DNA-U3 has multiple ORFs none are located with appropriately placed TATA boxes or polyadenylation signals, and are therefore unlikely to be transcribed (Burns *et al.*, 1995). However, a 10 kilodalton (kDA) protein has been detected in a DNA-U3 isolate (Beetham *et al.*, 1999). The function of DNA-U3 is still unknown. All six components were identified in a further eight isolates, each from a different country, suggesting that all six components are needed for infection (Karan *et al.*, 1994; Karan, 1997). These six components are consistently identified in infected plants and it is generally accepted that DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C, DNA-N are the integral components of BBTV (Figure 1.2).

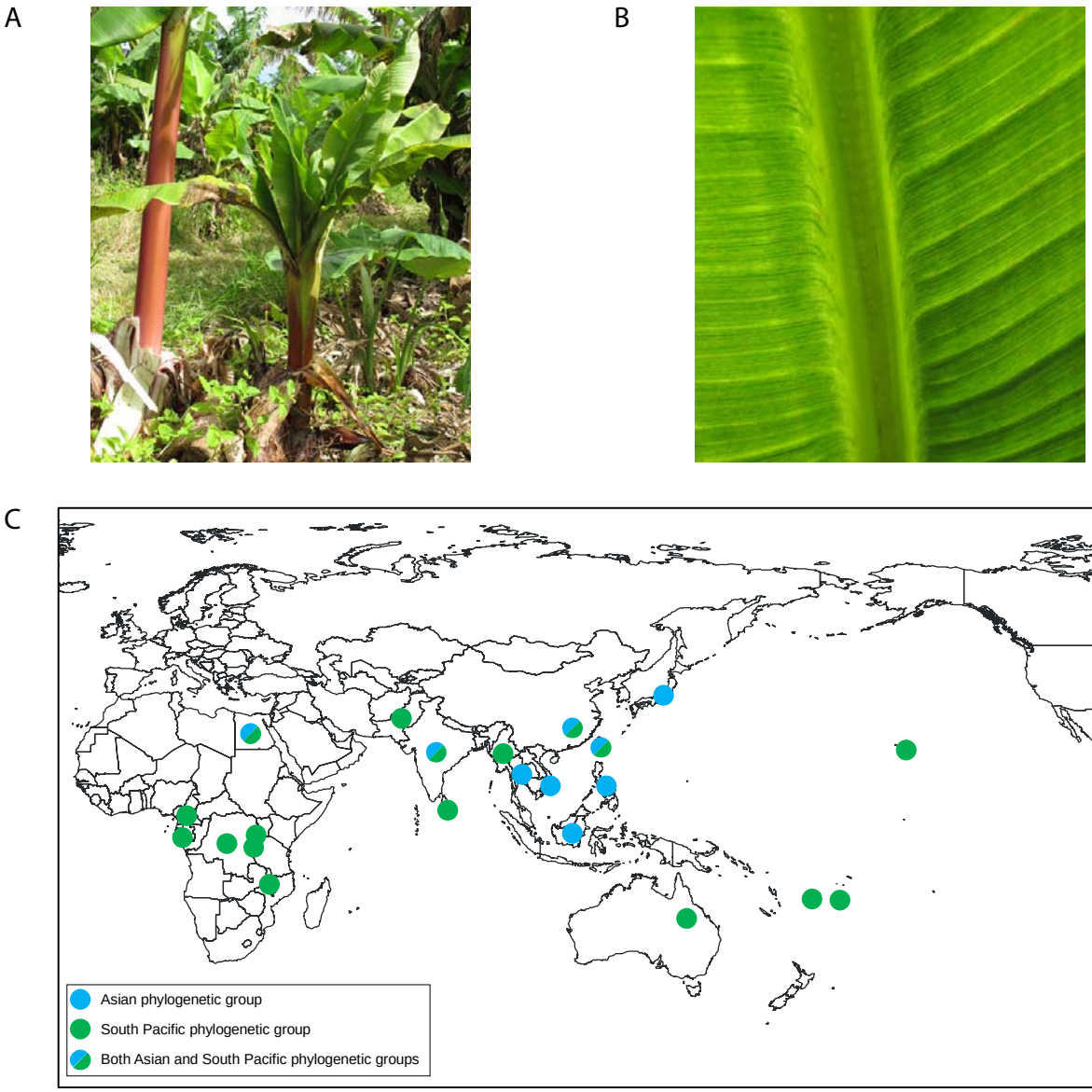


Figure 1.4: Symptoms of a BBTV infected plant, A) plant stunting and leaf bunching, B) hooking pattern on the midrib of the leaf. C) Distribution of BBTV isolates which have been sequenced (prior to the work described in this thesis), with associated references in Table 1.3. BBTV isolates fall into two phylogenetic groups, the South Pacific and Asian groups, with green or blue circles depicting the phylogenetic group of the isolates from each country. Blue/green circles depict countries with isolates from both phylogenetic groups.

Table 1.3: Countries with fully sequenced BBTV components (prior to work described in this thesis) and associated references.

Country	Sequence reference
Australia	(Burns *et al.*, 1995)
Burundi	(Wanitchakorn *et al.*, 2000b)
Democratic Republic of Congo	(Kumar *et al.*, 2011)
Cameroon	(Kumar *et al.*, 2011) and Unpublished
China	(He *et al.*, 2000; 2001a; b; Jun & Zhi-Xin, 2005; La *et al.*, 2000; Yu *et al.*, 2012) Unpublished
Egypt	(Abdel-Salam *et al.*, 2012; Karan *et al.*, 1994) Unpublished
Fiji	(Karan *et al.*, 1994; Wanitchakorn *et al.*, 2000b)
Gabon	(Kumar *et al.*, 2011)
Indonesia	(Furuya *et al.*, 2004; Pinili *et al.*, 2011)
India	(Anandhi *et al.*, 2007; Banerjee *et al.*, 2014; Islam *et al.*, 2010; Karan *et al.*, 1994; Selvarajan *et al.*, 2010; Vishnoi *et al.*, 2009) Unpublished
Japan	(Furuya *et al.*, 2005)
Sri Lanka	Unpublished
Myanmar	Unpublished
Malawi	(James, 2011; Kumar *et al.*, 2011)
Philippines	(Furuya *et al.*, 2006; Furuya *et al.*, 2005; Karan *et al.*, 1994; Wanitchakorn *et al.*, 2000b)
Pakistan	(Amin *et al.*, 2008; Bashir *et al.*, 2012; Hyder, 2009; Hyder *et al.*, 2007; Hyder *et al.*, 2011)
Rwanda	(James, 2011)
Thailand	Unpublished
Tonga	(Karan *et al.*, 1994)
Taiwan	(Fu *et al.*, 2009; Hu *et al.*, 2007; Karan *et al.*, 1994; Wanitchakorn *et al.*, 2000b)
United States of America	(Xie & Hu, 1995)
Vietnam	(Bell *et al.*, 2002; Furuya *et al.*, 2005; Karan *et al.*, 1994; Wanitchakorn *et al.*, 2000b)

1.4 BBTV

1.4.1 Phylogeny-based classification

Karan *et al.* (1994) identified two phylogenetic groups of BBTV based on the complete sequence of DNA-R, the South Pacific Group (SPG), containing isolates from Australia, Burundi, Egypt, Fiji, India, Tonga and Western Samoa and the Asian Group (AG), containing isolates from Taiwan, Vietnam and the Philippines. Sequences of the DNA-N isolates were also seen to fall into these two groups (Karan, 1997). As more isolate sequences have been determined all component sequences fall into these two geographical and phylogenetic groups, with only a few component which are found separately to those of the same region. As sequences from BBTV infected material from further countries have been identified it has been proposed that the two groups be renamed the Southeast Asia Group and the Pacific-Indian Ocean group (Yu *et al.*, 2012), however the majority of studies have continued to use the original group names.

1.4.2 Percentage pairwise identity

Percentage pairwise sequence identities are used to identify overall similarities between sequences and can also be used to identify the distribution of sequences for tentative strain differentiation which have been implemented in a number of ssDNA studies (Grigoras *et al.*, 2014; Muhire *et al.*, 2013; Stenzel *et al.*, 2014; Varsani *et al.*, 2014b). By aligning and comparing two sequences at a time, rather than creating a full alignment of all the sequences, the Sequence demarcation tool (SDT) determines a more accurate measure of the percentage pairwise identity between two sequences (Muhire *et al.*, 2014b). SDT can be used to identify similarities between sequences as well as identify the distribution of pairwise identities of the sequences. As BBTV is a multi-component genome, individual component pairwise identities can be compared as well as the full genome.

1.4.3 Insect transmission

BBTV is transmitted by the banana aphid, *P. nigronervosa*, in a persistent circulative manner. Early studies have shown that BBTV is unable to be transmitted by mechanical inoculations (Magee, 1927). *P. nigronervosa* require four hours of feeding on an infected plant for viral uptake but only 15 minutes for inoculation of a new plant, with longer feeding acquisition and transmission times increasing the transmission rates (Hu *et al.*, 1996). The juvenile

nymphs, however, require a much longer feeding time of at least 17 hours (Magee, 1940). After acquiring the virus a latent period of at least 20 hours is needed before the adult aphid is able to transmit the virus (Anhalt & Almeida, 2008). BBTV accumulates in the anterior midgut where it is thought to be internalised into the haemolymph of the aphid, before being transmitted to a new host through the principle salivary glands (Watanabe & Bressan, 2013; Watanabe *et al.*, 2013). BBTV is then able to be persistently transmitted for the duration of the life of the aphid (approximately 20 days) (Hu *et al.*, 1996). Replication studies of BBTV suggest the virus is not able to replicate in the vector (Hafner *et al.*, 1995).

As reviewed in Blanc *et al.* (2014), the specific molecular interactions between viruses and their vectors are still poorly understood, with the specific mechanisms of transfer across the vector tissues unknown. In geminiviruses, the CPs have been shown to be important in determining specific vector transmission (Briddon *et al.*, 1990). A viral helper factor, separate to the viral particles, has been suggested to be needed in order to transmit FBNYV; where the purified virus caused symptoms in the host, but was unable to be transmitted by the aphid vector (Franz *et al.*, 1999). This was also seen in a further study of FBNYV (Timchenko *et al.*, 2006) where the virus was unable to be transmitted from the symptomatic host with the aphid vector. However, a more recent nanovirus study successfully transmitted FBNSV with the insect vector (Grigoras *et al.*, 2009), suggesting that a helper factor may only be needed for some species.

Morphological and genetic differences, along with differences in host preference, has resulted in the proposed split of the aphid species *P. nigronervosa*, into two species; *P. caladii* and *P. nigronervosa* (Foottit *et al.*, 2010). *P. caladii* aphids prefer members of the Zingiberaceae and Araceae *and P. nigronervosa* prefer Musaceae (Foottit *et al.*, 2010). Regardless, infectivity tests of *P. caladii* from Taro, Ginger and Heliconia, have shown the proposed species is also able to experimentally transmit BBTV to banana plants (Watanabe *et al.*, 2013).

1.4.4 Alternative hosts of BBTV

Alternative hosts are important in disease control as they can act as viral reservoirs by maintaining the viral and insect population when the agriculturally important host is not present. An early study in India showed that BBTV was able to be passed from an infected banana plant, through a *Colocasia esculenta* which remained asymptomatic, into a healthy

banana which became visually infected (Ram & Summanwar, 1984). A further study in Taiwan used BBTV specific ELISA on two plant species *Canna indica* and *Hedychium coronarium* which showed mild to moderate reactions, suggesting these are also alternate hosts (Su *et al.*, 1992). Two BBTV isolates were recently successfully insect transmitted into three alternate hosts, in the case of an isolate from Japan *Alpinia zerumber*, *C. esculenta* and *C. indica* were infected (PCR verified) and an isolate from the Philippines infected *C. indica* (ELISA verified) (Pinili *et al.*, 2013). However, alternative host studies in Australia (Geering & Thomas, 1997), Hawaii (Hu *et al.*, 1996) and India (Manickam *et al.*, 2002) have failed to identify these species, and a number of further species (*Alocasia* spp., *Alpinia* spp., *Heliconia* spp., *Strelitzia* spp., *Xanthosoma* spp., *Zingiber officinale*, *Catheranthus roseus*), as alternative hosts. These studies, however, do not disprove the alternative hosts found previously, and it has been suggested that these results may be due to differences in the BBTV strains present, with only some strains able to infect alternative hosts (Geering & Thomas, 1997). ABTV is able to react with two of the monoclonal antibodies designed for BBTV (Sharman *et al.*, 2008), potentially suggesting that the weak/moderate ELISA results of Su *et al.* (1992) may also be reacting with an unknown member of the babuviruses.

1.4.5 Infectivity studies and component dynamics

Infectivity studies of BBTV have been unsuccessful and no alternative laboratory host has been identified. To date only one member of the *Nanoviridae*, FBNSV whose eight cloned components have been successfully used in infection studies and have remained insect transmissible (Grigoras *et al.*, 2009) suggesting that no more than eight components are needed. Previous to this, Timchenko *et al.* (2006) identified six components in FBNYV which were able to cause symptoms in a plant, however, these were not aphid transmissible.

1.4.6 Genome formula

The relative number of components for only one nanovirus, FBNSV, has been determined, revealing that the components are not present in equal numbers (Sicard *et al.*, 2013). Interestingly different genome formulas were identified for the same virus when it was introduced into the hosts *Vicia faba* and *Medicago truncatula* (viral isolate originally from *V. faba*). It has been suggested that the adjusting of the genome formula could be due to differences in the optimal level of the virus within the host, or that one host may be

maladaptive for the virus strain resulting in a less effective infection which presents as a different component formula (Sicard *et al.*, 2013).

1.4.7 Replication of nanoviruses

Neither geminiviruses nor nanoviruses encode a polymerase for replication therefore they utilise the hosts polymerase for replication. All circular Rep-encoding ssDNA (CRESS) viruses are thought to replicate through rolling circle replication (RCR), reviewed by (Gutierrez, 1999; Martin *et al.*, 2011; Timchenko *et al.*, 1999). Recombination-dependent replication (RDR) is a further, less studied, method of replication which has been identified in some geminiviruses (Alberter *et al.*, 2005; Jeske *et al.*, 2001; Preiss & Jeske, 2003).

1.4.7.1 Recombination-dependent replication

Large amounts of heterogenous length high molecular weight double stranded DNA (hDNA) has been identified in a number of geminiviruses which is thought to be the result of RDR (Jeske *et al.*, 2001; Preiss & Jeske, 2003), and has been reviewed briefly in Martin *et al.* (2011). Recombination-dependent replication (RDR) is thought to be a replication repair mechanism of ssDNA linear sequences which, either bind or invade a circular covalently closed DNA (cccDNA) at a site of sequence similarity. Once bound the short ssDNA is elongated with the cccDNA acting as a template. As the template cccDNA molecule is circular the resulting ssDNA can be much larger than a single molecule resulting in the hDNA identified in the geminiviruses (Alberter *et al.*, 2005; Jeske *et al.*, 2001; Martin *et al.*, 2011; Preiss & Jeske, 2003). It is still unknown whether RDR occurs in nanoviruses.

1.4.7.2 Rolling circle replication

The main mechanism for replication in the nanoviruses and geminiviruses is through rolling circle replication, reviewed in (Gronenborn, 2004; Gutierrez, 1999; Martin *et al.*, 2011). The Rep which is encoded on DNA-R is responsible for initiating replication all the integral components by first recognising and then nicking the nonanucleotide sequence (Hafner *et al.*, 1997b; Herrera-Valencia *et al.*, 2006). Briefly, component sequences first need to be converted into a double stranded replicative form which is primed by a component specific sequence which binds in the CR-M. Once Rep has recognised a component for replication it nicks the DNA at a specific recognition site in the CR-SL, creating a 3' OH overhang. This is identified by the host polymerase and replication commences around the circular molecule

(Hafner *et al.*, 1997b; Laufs *et al.*, 1995). The Rep molecule then joins the newly created ssDNA into a circular form (Hafner *et al.*, 1997b).

1.4.8 Common region major

The CR-M is present in all nanoviruses within the noncoding region of each component (Figure 1.2, Figure 1.5A). The CR-M of BBTV varies in size and location between components (Burns *et al.*, 1995; Vishnoi *et al.*, 2009). Each CR-M contains a GC rich region at the end, which shows similarities to promoter regions of geminiviruses (Burns *et al.*, 1995). However, neither the CR-M or the CR-SL are essential for promoter activity, only the deletion of the intergenic region containing the TATA box resulted in the absence of promoter activity (Dugdale *et al.*, 1998). The CR-M is also involved in the priming of the complementary strand of the ssDNA molecule into the double-stranded replicative form, with primers which bind across the noncoding region and start of the CR-M (Hafner *et al.*, 1997a). These approximately 80 bp primers are component specific and appear to be packaged within each virion, with primers identified which bind to DNA-R, DNA-U3, DNA-S, DNA-M and DNA-C, with DNA-S primer thought to be present but at undetectable concentrations (Hafner *et al.*, 1997a). Interestingly DNA-C, which had yet to be named or be assigned a function, had a primer which was found in much higher concentrations than the other primers suggesting its importance early in infection (Hafner *et al.*, 1997a). It is now known that DNA-C is involved in increasing replication through interacting with the host's cell cycle (Aronson *et al.*, 2000; Lageix *et al.*, 2007; Wanitchakorn *et al.*, 2000a).

Once expressed Rep is required to recognise components for replication through repeat sequences also known as iterons which are present in the CR-SL of the integral components. The CR-SL also contains the recognition sequence for the nick site for initiation of replication.

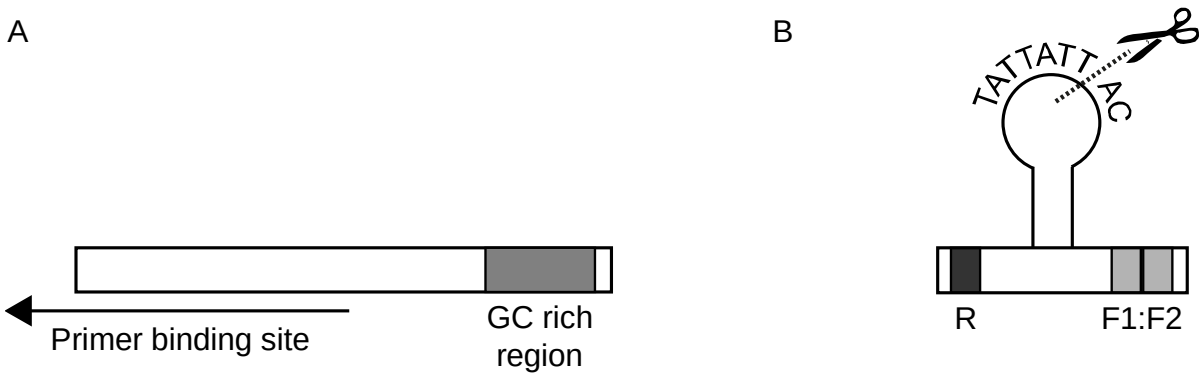


Figure 1.5: Cartoon representation of BBTV common regions A) The common region major with the GC rich region and the primer binding site for secondary strand synthesis. B) The common region stem-loop with the nonanucleotide sequence which is nicked by the replication-associated protein. The three iterons; R, F1 and F2 are shown relative to the stemloop region.

1.4.9 Common region stem-loop

As the nanoviruses are multi-component, all components of a species contain a conserved region, the CR-SL, which is conserved across all components in a species to enable the rep to recognise and replicate the associated viral components. The CR-SL contains three iteron sequences, iteron R (GTCCC) upstream of the stem-loop, iteron F1 (GGGAC) and iteron F2 (GGGAC) downstream of the stem-loop, with the majority of babuvirus components containing these three iterons (Figure 1.5B) (Burns *et al.*, 1995; Mandal *et al.*, 2013; Sharman *et al.*, 2008). Complete mutation of each of the three iterons in BBTV DNA-N was shown to reduce replication by an intact DNA-R, suggesting all three are involved in Rep recognition (Herrera-Valencia *et al.*, 2006). However, the mutation of iterons resulted in different levels of replication reduction with iteron F2 likely the most crucial recognition iteron with replication reduced to barely detectable levels when mutated, followed by iteron R which showed nearly 62% reduction and iteron F1 with nearly 42% reduction in replication (Herrera-Valencia *et al.*, 2006). Interestingly, although both iteron F1 and iteron F2 have identical sequences and they are located adjacent to one another, there is a large difference in the replication reduction when mutated.

In CBDV DNA-N a short CR-SL has been identified which does not contain the iteron R (Mandal *et al.*, 2013). BBTV DNA-U3 has also been suggested to have a short CR-SL without an iteron R (Banerjee *et al.*, 2014; Burns *et al.*, 1995; Vishnoi *et al.*, 2009), however Herrera-Valencia *et al.* (2006) has identified a potential iteron R sequence much further upstream of the stem-loop of DNA-U3. Since the mutation of iteron R lowers the replication of the components, but does not halt replication it is possible that these components are able to exist without an iteron R. However, all other babuviruses contain iteron R, plus discrepancies are seen with BBTV DNA-U3, and only a single isolate was available for CBDV, therefore further studies may elucidate the presence of the iteron R further upstream for both these components.

Interestingly MDV, FBNYV and SCSV, all of which infect legumes, have similar iterons and Reps, and are able to replicate the DNA-S of the other two species, suggesting potential reassortment across species (Timchenko *et al.*, 2000). Although this has yet to be tested in babuviruses, similarities in the iterons of babuviruses suggest that this may also be possible for the three babuviruses species, in particular BBTV and ABTV, as both are able to infect abaca and banana.

The CR-SL also contains a stem-loop sequence with a highly conserved nine nucleotide motif, the nonanucleotide NANTATTAC, which is present in all CRESS viruses, and is located in the loop section of the stem-loop (Rosario *et al.*, 2012). The nonanucleotide of the babuviruses is TATTATTAC and nanoviruses TAGTATTAC (Rosario *et al.*, 2012). Once a component has been recognised, the Rep initiates replication by creating a nick in the stem-loop at a specific site. Like the geminiviruses the Rep of BBTV nicks the nonanucleotide motif between the 7th and 8th nucleotide TATTATT↓AC in the virion sense strand resulting in an OH' overhang (Hafner *et al.*, 1997b; Laufs *et al.*, 1995). This overhang is recognised by the host polymerase which initiates elongation of the DNA. Interestingly the removal of the stem section of the stem loop from either side of the nonanucleotide, or mutation of the stem sequence still resulted in the Rep cutting the motif at the same site TATTATT↓AC suggesting it is the nonanucleotide sequence itself rather than the stem which is important in the nicking activity (Hafner *et al.*, 1997b). Once the host polymerase has finished replication the Rep joins the ends of the displaced original virion producing a circular molecule (Hafner *et al.*, 1997b; Laufs *et al.*, 1995).

1.4.10.1 Motifs

1.4.10.2 Rep Motifs

The Rep contains six highly conserved motifs which are all involved in replication and are seen in all CRESS viruses, RCR I, RCR II, RCR III, and the SF3 Helicase Motifs; Walker A, Walker B and Motif C (Figure 1.6A and 6C) (Gorbalenya *et al.*, 1990; Rosario *et al.*, 2012; Vega-Rocha *et al.*, 2007). As reviewed in Rosario *et al.* (2012) these three RCR motifs are thought to be involved in recognition and nicking of the DNA for replication.

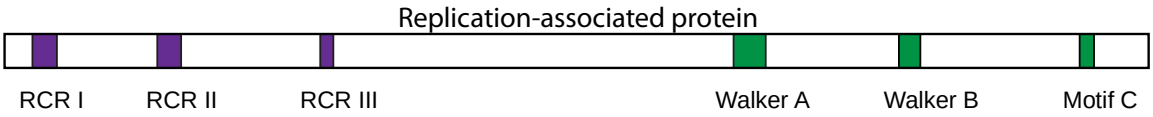
RCR I is involved in Rep recognition and is associated with the specificity determinates (SPDs) located either side of the RCR I and RCR II motif (Argüello-Astorga & Ruiz-Medrano, 2001; Londoño *et al.*, 2010). The SPDs are further Rep recognition sites directly related to the iteron sequences seen in the CR-SL. Geminiviruses have a number of different SPDs across species suggesting species with different SPDs are unable or less able to co-replicate. However BBTV and ABTV have identical SPD_{r1} (RYVVC) and SPD_{r2} (RSS) sequences, as do MVD, FBNYV and SCSV with SPD_{r1} (RQVIC) and SPD_{r2} (RTS)(Londoño *et al.*, 2010). Interestingly, as mentioned earlier these nanoviruses have been shown

experimentally to co-replicate (Timchenko *et al.*, 2000) suggesting BBTV and ABTV may also co-replicate.

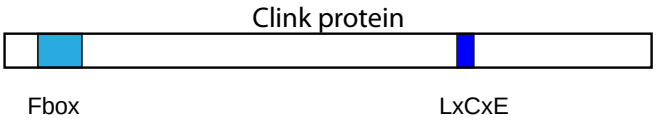
A study on *Tomato yellow leaf curl virus*, a geminivirus, has shown RCR II motif is involved in binding the Mg^{2+} or Mn^{2+} cations needed to catalyse the nicking of the nonanucleotide (Laufs *et al.*, 1995). BBTV also requires Mg^{2+} or Mn^{2+} to catalyse the nicking reaction, however Ca^{2+} is also able to act as a catalyst, although at a much lower rate (Hafner *et al.*, 1997b). The RCR III motif is the active site of cleavage of the nonanucleotide and the joining of the replicated ssDNA into a circular form (Hafner *et al.*, 1997b; Laufs *et al.*, 1995; Orozco & Hanley-Bowdoin, 1998).

These three RCR motifs are located in structurally equivalent positions in the Rep proteins of the nanoviruses, geminiviruses and circoviruses (Vega-Rocha *et al.*, 2007). The region containing the Walker A, Walker B and Motif C has been shown in geminiviruses to be a helicase domain which is involved in the unwinding of the dsDNA in a 3'-5' direction for leading strand synthesis (Choudhury *et al.*, 2006; Clérot & Bernardi, 2006). Based on similarities in motifs and functions of the nanovirus Rep, the Walker A, Walker B and Motif C are also expected to be helicase motifs.

A



B



C

	RCR I	RCR II	RCR III	Walker A	Walker B	Motif C	F-box	LxCxE
BBTV	MFTINN (VFTINN)[1] (MFTIND)[1]	RHVQG	YCMK (CCMK)[1] (YCTK)[1]	GPNGGEGKT (GPNEGEGKP)[1]	VIFDI (CIFDI)1 (VILDI)[2]	VMAN (LMAN)[1] (VIVN)[1]	MPDDVKREIKEI (MPDDVKRVIKEI)[30] (MPDDVKKEIKEI)[1] (MSDDVKREIKEI)[1] (MPDDVKRVIKEK)[2] (IPDDVKRVIKEI)[1] (MTDHVKREIKEI)[1] (MPDDVEREIKEI)[1] (MPDDVEREIKEI)[1]	LYCDE (LFCDE)[3] (LYCDA)[2] (FYCDE)[1]
ABTV	MFTINN	RHVQG	YCMK	GPNGGEGKT	VIFDI	VMAN	FCDDVKRVIKQK	LFCDE
CBDV	MFTINN	RHVQG (RNVQG)[1]	YCMK	GPNGGEGKT	VIFDI	VMAN	VYPDDVKRVIKEYY (VYPDDIKRVIKEYY)[5] (VSPDDVKRVIKEYY)[2] (VYPDDLKRVIKEYY)[1] (VYPDDVKRAIKEYY)[1] (VYPVDVKRVIKEYY)[1]	LFCDE

Figure 1.6: Illustration of the location of the babuvirus conserved motifs on A) the replication-associated protein, with the three rolling circle replication (RCR) motifs and the helicase motifs, walker A, walker B and motif C, B) the Clink protein with the F-box and LxCxE motifs. C) The babuvirus conserved motif sequences, were identified in all publically available babuvirus sequences, based on Clink motifs in Aronson *et al.* (2000) and DNA-R motifs as reviewed in Rosario *et al.* (2012). Alternative motifs are shown in brackets with the number of sequences which contain the alternative motif in superscript.

1.4.10.3 Clink Motifs

The Clink, which has only been identified in nanoviruses, also contains conserved motifs, the F-box motif and the LxCxE motif located 9-20 amino acids (aa) and 110/111-114/115 aa respectively from the start of the protein (Figure 1.6B) (Aronson *et al.*, 2000). The LxCxE motif is involved in binding to the plant retinoblastoma protein and switching the plant into replication or S-phase (Aronson *et al.*, 2000; Lageix *et al.*, 2007; Wanitchakorn *et al.*, 2000a). Binding of retinoblastoma proteins to increase viral replication has also been identified in mammalian DNA viruses, although through a different motif than the LxCxE (Chellappan *et al.*, 1992). Studies of FBNYV have identified that the Rep is able to replicate without the Clink, however when the Clink is present replication increases 3-7 fold (Aronson *et al.*, 2000). This increase can be directly contributed to the LxCxE, as mutation in this motif removes this replication increase (Aronson *et al.*, 2000; Hipp *et al.*, 2014). Therefore to increase replication of the viral components the Clink encoded on DNA-C, is likely to be expressed early in replication along with the Rep encoded on DNA-R. Some geminiviruses contain an LxCxE motif, however it is present on the Rep rather than on a separate protein, therefore it is directly linked to Rep expression (Xie *et al.*, 1995). The other motif present on the Clink, the F-box motif, does not directly influence replication as mutation has no effect on the replication efficiency (Aronson *et al.*, 2000). The F-box motif binds to a protein (SKP1) in the SCF complex, which is involved in the ubiquitination pathway and it has been suggested that the F-box may be utilised later in replication as nanoviruses do not cause permanent cell proliferation (Aronson *et al.*, 2000). Although no F-box motif has been identified in the geminiviruses, the *trap* is also able to interact (indirectly) with the SCF complex (Lozano-Durán *et al.*, 2011).

Comparison of the Rep motifs of BBTV, ABTV, CBDV show motifs are highly conserved within species and within the babuviruses (Figure 1.6C). If the small number of isolates with divergent Rep motifs are excluded (motifs in brackets in Figure 1.6C), all Rep motifs across the three babuviruses are identical. The F-box motif shows much greater diversity within species as well as across babuviruses, however, five amino acids within the F-box DDVKR are found in the majority of all motifs suggesting these are the key amino acids in this motif.

1.4.10 Evolution of nanoviruses

The multi-component genome of BBTV is able to utilise random mutation, recombination and reassortment to increase its genetic diversity and evolve. Recombination and reassortment facilitate rapid genome changes through replacement of large sections of the viral genome. Substitution rates of other ssDNA viruses and nanoviruses suggest these viruses are evolving faster than previously thought (Duffy *et al.*, 2008). Identifying patterns of selection in ORFs as well as reassortment and recombination patterns allows a greater understanding of how this virus is evolving.

1.4.11.1 Substitution rates

Due to DNA viruses utilising the error-repairing host polymerases for replication it was initially thought that these viruses would have lower mutation and substitution rates than the RNA viruses which use error prone polymerases. However, it appears that more than just the host polymerase effects the rates of substitution, with ssDNA viruses evolving more rapidly than previously thought, with similar substitution rates per site per year to those of similar sized RNA viruses (Duffy *et al.*, 2008). It has been suggested that the high mutation and substitution rates could be caused by the host polymerase not acting as a proof reading polymerase due to methylation differences of the viral DNA, or due to viral interference of the polymerase (Arguello-Astorga *et al.*, 2007; Duffy *et al.*, 2008). Substitution rates have been calculated for a number of ssDNA viruses (De Bruyn *et al.*, 2012; Duffy & Holmes, 2008; 2009; Firth *et al.*, 2009; Grigoras *et al.*, 2010b; Harkins *et al.*, 2014; Harkins *et al.*, 2009; Kraberger *et al.*, 2013; Shackelton *et al.*, 2005; Streck *et al.*, 2011) with a member of the nanoviruses, FBNSV, among the fastest known ssDNA virus with a substitution rate of $1.78x10^{-3}$ substitutions per site per year (Grigoras *et al.*, 2010b). Patterns of selection within ORFs are also being used to identify specific regions which are evolving under positive and negative selection (Kraberger *et al.*, 2015; Muhire *et al.*, 2014a; Stenzel *et al.*, 2014).

1.4.11.2 Recombination

Recombination is a mechanism of increasing genetic diversity. Rather than a single point mutation, recombination allows fragments to be 'acquired' by homologous recombination resulting in faster genetic divergence. Recombination of ssDNA viruses in eukaryotes has been reviewed by Martin *et al.* (2011). The ssDNA viruses are able to utilise recombination to increase genetic diversity and experimental observations of chimeric geminiviruses, MSV

sequences, are able to find an optimal fitness landscape through recombination (Monjane *et al.*, 2014; Monjane *et al.*, 2012; van der Walt *et al.*, 2009). Recombination in multi-component viruses can be both, intra-component (between two of the same component type), or inter-component (between different component types). Recombination in multi-component viruses across the common regions has been suggested as an important mechanism for maintaining control of all components (Hughes, 2004). In order for recombination to occur multiple viruses or virions are required within the same cell.

Recombination events which result in a disrupted ORF/s are unlikely to be successful in infection and hence recombination breakpoints are more frequently identified in the intergenic regions than within the coding regions (Lefeuvre *et al.*, 2009). A number of defective molecules have been identified which are likely the result of recombination or deletion events. Defective genomes have been identified associated with a number of viruses (Bach & Jeske, 2014; Casado *et al.*, 2004; Frischmuth & Stanley, 1992; Hadfield *et al.*, 2012; Horn *et al.*, 2011; Paprotka *et al.*, 2010a; Patil & Dasgupta, 2006; Stanley & Townsend, 1985; Stenger *et al.*, 1992; van der Walt *et al.*, 2009; Zaffalon *et al.*, 2012) including a number of defective genomes associated with BBTV in GenBank. These genomes are recognisable as the component they arose from but are defective due to deletions and/or insertions and are no longer functional usually due to a disrupted ORF or due to a disruption in a region required for replication such as a common region. Although defective in the sense that they are no longer functional, these sequences, along with the alphasatellites are important sources of genetic diversity which may be utilised during recombination.

Multiple mechanisms of recombination have been theorised, although it is currently unclear exactly how recombination is occurring. One potential mechanism is displacement of a replicating strand due to replication machinery clashes during concurrent replication and transcription, followed by reattachment to a different template strand and completion of replication (Owor *et al.*, 2007a). It is also possible that breaks in ssDNA during replication may be repaired through RDR utilising a different template (overview in replication section) (Jeske *et al.*, 2001; Preiss & Jeske, 2003). Host repair mechanisms may also be involved in recognising and repairing breaks in double stranded viral DNA. Potentially a combination of all of these mechanisms may be contributing to the recombination which is being detected in ssDNA genomes.

Recombination has been detected in a number of geminiviruses (Jovel *et al.*, 2007; Kraberger *et al.*, 2013; Saunders *et al.*, 2002) and other ssDNA viruses (Julian *et al.*, 2013; Rez *et al.*, 2014; Stenzel *et al.*, 2014). Recombination analyses of nanovirus genomes identified 23 recombination events, with at least one recombinant from BMLRV, FBNYV, FBYLV, FBNSV, PYSV, PNYDV, and MDV species (Grigoras *et al.*, 2014). Both inter- and intra species recombination events were detection with all species showing evidence of inter-species events (Grigoras *et al.*, 2014). DNA-U2 was the most recombinant component with six events detected (Grigoras *et al.*, 2014). Recombination has also been identified in CBDV, with both inter- and intra-component recombination events detected (Savory & Ramakrishnan, 2014a). Although very few sequences are available for ABTV, recombination events have also been detected in ABTV, as reported in Chapters Two and Three.

Large numbers of recombination events have been detected in BBTV components. An early recombination study identified phylogenetically that the CR-M of DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N from the SPG fall together rather than with the corresponding component in the AG (and vice versa) which is likely to be due to recombination across the components within the two groups (Hu *et al.*, 2007). Specific recombination events involving the CR-M have also been detected with a DNA-U3 isolate containing a DNA-N-like CR-M (Islam *et al.*, 2010) and a DNA-U3 component containing a CR-M from a satellite genome (Fu *et al.*, 2009). Multiple BBTV recombination events involving the CR-M have been detected and are highlighted in Chapter Two, with a recombination hotspot also detected at the CR-M.

Recombination has also been detected across the BBTV genome, in regions other than the CR-M (Banerjee *et al.*, 2014; Hyder *et al.*, 2011; Wang *et al.*, 2013). By analysing all worldwide BBTV isolates available in 2011 and 2014, along with those sequenced as part of this thesis, the number of detected recombination events has greatly increased, suggesting recombination is an important mechanism in the evolution of BBTV (described in Chapters Two and Three).

1.4.11.3 Reassortment

Reassortment, also known as pseudo-recombination, is the swapping of entire segments or components and is a further mechanism which multi-component and multi-segmented viruses are able to utilise to increase the genetic diversity of the viral genome. Like recombination,

reassortment requires multiple virions present to facilitate the swapping of genetic material. Reassortment in influenza has been intensively studied with a number of reassortants resulting in highly pathogenic strains (Creanga *et al.*, 2013; Guan *et al.*, 2002; Nelson *et al.*, 2008; To *et al.*, 2013). Segmented RNA plant-infecting viruses also show evidence of reassortment (Chen *et al.*, 2007; Gu *et al.*, 2007; Maoka *et al.*, 2010). Interesting reassortment events involving the RNA virus, *Tomato spotted wilt virus*, have resulted in infections of previously resistant tomato and pepper cultivars (Qiu & Moyer, 1999; Tentchev *et al.*, 2011).

Genome reassortment has also been identified in the multi-component geminiviruses (Chen *et al.*, 2009; Idris & Brown, 2004; Pita *et al.*, 2001) and in a number of nanovirus genomes (Grigoras *et al.*, 2014; Hu *et al.*, 2007; Savory & Ramakrishnan, 2014a; Yu *et al.*, 2012). Nanovirus and multi-component geminivirus genomes are not only segmented, but each segment is also encapsidated separately. As there are no physical ties between the components, these viruses are able to reassort not only when they are uncapsidated, but also when they are encapsidated during movement around the host or within the insect vector.

In order to increase the likelihood of detecting reassortment events, full genomes rather than disparate components, are preferable. A recent reassortment analysis of 163 full genomes of CBDV identified 69 genomes which contained at least one reassortment event (Savory & Ramakrishnan, 2014a). All six integral components (DNA-Uf1 and DNA-Uf2 were not identified in the study) were transferred in at least one reassortment event, with two components, DNA-M and DNA-N, the most commonly transferred components (Savory & Ramakrishnan, 2014a).

Reassortment has been identified in the genomes; FBNSV (n=3 genomes), FBNYV (n=5 genomes), BMLRV (n=1 genome) and MDV (n=1 genome) (Grigoras *et al.*, 2014). Although only a small dataset was available 12 reassortment events were detected in ten genomes. Interestingly the most commonly transferred component in FBNSV was DNA-N (two events) and in FBNYV was DNA-M (three events) (Grigoras *et al.*, 2014) which, although low numbers, were similar to CBDV. This suggests that these components are either more likely to reassort than others or that these reassortants are not selected against. In order for reassortment events to be successful, the reassorted component needs to not only carry out its role in relation to the other components, but also be recognised and replicated by the Rep. The numbers of reassortment events detected in the nanovirus genomes, as well as multiple

reassortment events within a small number of genomes of FBNSV, FBNYV and CBDV, suggest that reassortment is prevalent.

Previously only three BBTV studies have identified potential reassortment events, both by identifying phylogenetic anomalies (Hu *et al.*, 2007; Yu *et al.*, 2012). Hu *et al.* (2007) identified a reassorted isolate by comparing the phylogenetic group placing, as three components were present in the SPG, and one (DNA-M) in the AG. Yu *et al.* (2012), used a more generalised approach noting that between two components DNA-R and DNA-S there were changes in the clade grouping of isolates within the AG. At the time of either study there were very few components available with 102 and 125 sequences analysed in (Hu *et al.*, 2007) and (Yu *et al.*, 2012) respectively. The increase in BBTV components and full genomes allowed the detection of eight reassortment events (Chapter Two), followed by 40 detected reassortment events (Chapter Three). Therefore reassortment in BBTV appears to be a frequent occurrence.

1.5 Banana-infecting badnaviruses

Banana streak disease is caused by a pararetrovirus, (family *Caulimoviridae*, genus *Badnavirus*) which infects *Musa* species. Members of the *Caulimoviridae* are the only dsDNA viruses which are known to infect plants. There are seven different genera the *Caulimovirus*, *Cavemovirus*, *Solendovirus*, *Petuvirus*, and *Soymovirus* which have isometric particles and the *Badnavirus* and *Tungrovirus* which have bacilliform shaped particles. The dsDNA is open-circular and ranges in size from 7.2-9.2 kb. A further Florendovirus genus, has recently been proposed which consists of endogenous sequences detected in a number of flowering plant species (Geering *et al.*, 2014). No episomal florendovirus sequences have been detected, however, reconstructions of the endogenous forms as well as entire copies of some species have been detected within plant genomes, interestingly some sequences suggest florendoviruses may have bipartite genomes (Geering *et al.*, 2014).

The *Badnavirus* genus is the largest genus of the *Caulimoviridae* and consists of 25 accepted species which infect both monocotyledonous and dicotyledonous plants with the type species *Commelina yellow mottle virus*. Badnavirus species are determined based on >20% difference in the nucleotide sequence of the polymerase (Reverse Transcriptase (RT) and ribonuclease H (RNAse H)), host ranges, gene sequences and vector specificities (King *et al.*, 2011). There are four accepted species of BIB; BSGFV (Gayral *et al.*, 2008), BSMYV (Geering *et al.*,

2005b), BSOLV (Harper & Hull, 1998) and BSVNV (Lheureux *et al.*, 2007) An additional six BIB are yet to be accepted by ICTV, Banana streak IM virus (BSIMV), Banana streak CA virus (BSCAV), Banana streak UA virus (BSUAV), Banana streak UI virus (BSUIV), Banana streak UL virus (BSULV), Banana streak UM virus (BSUMV) (Geering *et al.*, 2011; James *et al.*, 2011b). BIB are transmitted by several species of mealybug, including *Planococcus citri*, *Dysmicoccus brevipes* and *Planococcus ficus* (Geering *et al.*, 2005b; Lheureux *et al.*, 2007; Meyer *et al.*, 2008).

BSGFV, BSMYV, BSOLV and BSIMV are named after the banana cultivars they were first identified from, Goldfinger, Mysore, Obino l'Ewai and Imové, respectively. BSOLV was originally Banana streak ON virus due to it being identified in Onne station in Nigeria (Harper & Hull, 1998), but was also identified in Obino l'Ewai and named BSOLV (Geering *et al.*, 2001). BSVNV was first found in Vietnam and BSUAV, BSUIV, BSULV, BSUMV are isolates A, I, L, M from Uganda.

1.5.1 Genome

Badnaviruses have dsDNA genomes with bacilliform particles which are 30 x 130 nm in size (Lockhart, 1990; Lockhart, 1968). The ~7.5 kb genome of BIB contains three ORFs (Figure 1.7), with ORF 1 encoding a protein of unknown function. ORF 2 encodes the virion associated protein (Stavolone *et al.*, 2001). ORF 3, the largest ORF encodes a 208 kDa polyprotein which contain domains suggesting this ORF is proteolytically cleaved to produce the Mp, CP, aspartic protease, RT and RNase H (Geering *et al.*, 2005b; Harper & Hull, 1998).

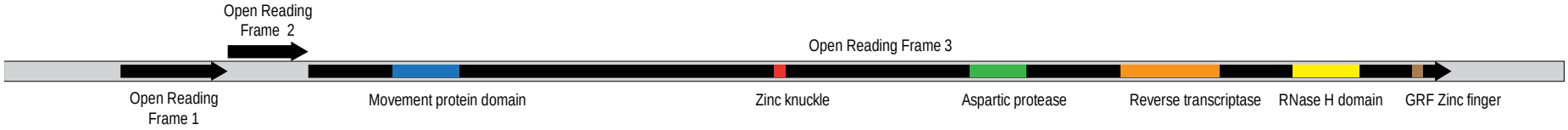


Figure 1.7: Cartoon representation of a linearised episomal banana-infecting badnavirus genome. The genome is linearised at the 5' of the $tRNA^{met}$ binding site which is the convention of plant pararetroviruses. ORF 2 overlaps the end of ORF 1 and the start of ORF 3. The domains of the polyprotein ORF 3, as identified in Pfam, are shown.

1.5.2 Replication

Caulimoviruses are pararetroviruses which replicate through an RNA intermediate, but unlike retroviruses they do not contain an integrase and replication is episomal, occurring in the nucleus rather than within the host DNA (King *et al.*, 2011). The circular genome of all caulimoviruses is discontinuous, with the ssDNA genome breaks first repaired in the nucleus before the genome is transcribed by the host polymerase. This host transcribed RNA then acts as a template for both translation of the ORFs and reverse transcription to form the minus strand DNA genome. Reverse transcription is carried out by the viral encoded RT.

It is thought that the integration of badnaviruses (endogenous form) into the host genome is through recombination during the repair of these ssDNA breaks, or other replication breaks (Gayral *et al.*, 2008; Geering *et al.*, 2014; Hohn *et al.*, 2008). These endogenous forms are replicated along with the host genome.

1.5.3 Endogenous form

Caulimoviruses have been identified integrated in the host genome, with the vast majority of integration sequences fragmented, rearranged and containing mutations, reviewed in Teycheney and Geering (2011). Caulimoviruses do not contain an integrase gene and therefore it is thought that integration events of the viral genome into that of the hosts is likely due to recombination events during replication or the repair of DNA breaks (Gayral *et al.*, 2008; Geering *et al.*, 2014; Hohn *et al.*, 2008).

A number of endogenous banana-infecting badnavirus (eBIB) genomes have been detected in the genomes of *M. balbisiana*, *M. acuminata* and *M. schizocarpa* (Gayral *et al.*, 2008; Geering *et al.*, 2001; Geering *et al.*, 2005a; Harper & Hull, 1998; Harper *et al.*, 1999; Ndowora *et al.*, 1999). These eBIB show a large amount of variation, suggesting a number of different badnavirus integration events (Geering *et al.*, 2005a).

1.5.4 Activatable endogenous viruses

Although the majority of the endogenous forms of caulimoviruses are thought to be replication-defective, a few species have been seen to cause infection including *Tobacco vein clearing virus* (*Solendovirus* genus) (Lockhart *et al.*, 2000) and *Petunia vein clearing virus* (*Petuvirus* genus) (Richert-Pöggeler *et al.*, 2003). Isolates of the endogenous BSOLV, BSIMV and BSGFV are able to revert to the episomal form and cause infection. These

infective endogenous forms are prefixed with "ea" for endogenous and activatable (Staginnus *et al.*, 2009). An eaBSOLV was first identified when tissue culture progeny of the cultivar Obino l'Ewai, developed banana streak disease in the absence of the disease in the parental plants, further investigation revealed BSOLV within the *Musa* genome (Harper *et al.*, 1999; Ndowora *et al.*, 1999). It was later confirmed that tissue culture is able to activate eBIB (Côte *et al.*, 2010). EaBSOLV is present in the *Musa* genome in two regions which are separated by a large section (6 kb) of partial viral sequences, including inversions and deletions, which is thought to be excised during a recombination event (Chabannes *et al.*, 2013; Harper *et al.*, 1999; Ndowora *et al.*, 1999). None of the currently identified eaBIB genomes are present in the *Musa* genome in an intact form with both eaBSGFV and eaBSIMV also within a large viral section with viral inversions and deletions requiring recombination in order to produce full length genomes (Chabannes *et al.*, 2013; Gayral *et al.*, 2008; Iskra-Caruana *et al.*, 2010). Both eaBSGFV and eaBSOLV regions are located on chromosome 1 and eBSIMV on chromosome 2 (Chabannes *et al.*, 2013). These eaBIB sequences have only been identified in the B genome and are an issue for banana production as they are likely present in the banana breeding stocks, with all recent worldwide Banana streak disease emergences thought to be due to eaBIB species (Chabannes *et al.*, 2013).

1.6 Strategies for development of virus resistant bananas

A number of strategies exist for producing resistant crops such as the expression of mutated viral proteins, peptide aptamers, RNA interference (RNAi), and virus induced cell death, as reviewed in Elayabalan *et al.* (2015). A number of studies have produced resistant banana cultivars through Rep targeted RNAi, with the transgenic bananas resistant for the six month testing period (Borth *et al.*, 2011; Elayabalan *et al.*, 2013; Shekhawat *et al.*, 2012). However an earlier siRNA study identified delayed, rather than total resistance to BBTV, with symptoms delayed for up to a year before the plants developed BBTV. This was thought to be due to chimeric plants but may suggest longer resistance testing periods are needed (Borth *et al.*, 2009).

A further resistance strategy is the in-plant activation (INPACT) system, which uses an integrated split gene cassette which is only released in the presence of the virus, resulting in inducible resistance (Dugdale *et al.*, 2013). The INPACT cassette contains the viral stem-loop region and is released through RCR by the target virus to form a circular episomal INPACT cassette. The episomal cassette can then continue to be replicated through RCR,

and/or transcribed and translated to produce the lethal protein also present on the cassette, killing the virus infected cell (Dugdale *et al.*, 2013). Therefore the INPACT system can be used to target viruses which replicate though RCR such as geminiviruses and nanoviruses, with an INPACT based platform recently used in MSV infected cell-culture resulting in an inducible inhibition of MSV (Shepherd *et al.*, 2014).

As the spread of BIB via mealybugs has been reported as slow (Daniells *et al.*, 2001), and the greatest source of infections is likely due to eaBIB species (Chabannes *et al.*, 2013), the identification of virus free breeding stock will potentially decrease the incidence of banana streak disease without genetic modification.

1.7 Methods for detection of BBTV and BIB

1.7.1 Polymerase chain reaction

Polymerase chain reaction (PCR) is a well utilised technique for identifying known viruses from samples and is routinely used for screening purposes for both BBTV and BIB. Back-to-back (degenerate primers) can also be used to identify the full continuous sequence of circular viruses. Full genome sequencing of BIB is rare, with primers which amplify the RNase H region often used for species determination, rather than the entire genome. Although, screening for a region of BIB is unable to differentiate between an active episomal infection or an endogenous sequence, it still allows for the detection of unclean banana stock, as the endogenous sequence is a potential reservoir. BBTV components are relatively small, with a number of components sequenced from around the world, however, the vast majority of these are disparate components with no full genomes available.

1.7.2 Enzyme-linked immunosorbent assays

ELISA is based on antibodies recognising the CP of BIB, therefore they only detect the episomal form and not the endogenous forms. However, since banana streak disease is caused by a number of badnaviruses and even triple antibody (TAS) ELISA have been seen to cross-react with closely related viruses, *Rice tungro bacilliform virus* and *Sugarcane bacilliform virus* (ScBV) (Manoranjitham *et al.*, 2012). Although sensitive at detecting infection, when viral titres are low such as in asymptomatic plants, TAS-ELISAs can produce false negatives (Thottappilly *et al.*, 1998). Interestingly BIB infected plants can fluctuate in symptom

expression and detectible viral levels, resulting in changes in ELISA detection of these species (Dahal *et al.*, 1998).

BBTV has a much lower serological diversity, with ELISA a routinely used diagnostic tool. Two of the ten antibodies used in BBTV ELISA cross-react with ABTV and therefore these can also be used to identify ABTV infected *Musa* plants (Sharman *et al.*, 2008).

1.7.3 Imunocapture PCR

Immunocapture PCR is able to combine the antibody specificity of ELISA with sequence specificity of PCR in order to amplify and sequence only the episomal form and hence is often used for badnaviruses. Badnavirus specific antibodies are used to bind the CP of the episomal form to a PCR tube (Geering *et al.*, 2000; Le Provost *et al.*, 2006). After the episomal forms are bound, washing removes any unbound substances including plant DNA, with positive BIB specific PCR attributed to episomal forms rather than integrated forms. However, isolates with diverse coat proteins, and/or diverse sequence, may still not be detected, as they either do not bind to the antibodies or are too divergent for the PCR primers.

1.7.4 Enrichment of circular molecules and PCR

The rolling circle amplification (RCA) technique has revolutionised the study of circular DNA viruses (Owor *et al.*, 2007b; Shepherd *et al.*, 2008). Phi29 is a polymerase which replicates DNA in a strand displacement manner, resulting in concatemers of DNA rather than single molecules. Due to strand displacement activity Phi29 preferentially amplifies circular DNA. Multiple polymerases are able to bind to each DNA molecule further enhancing replication. Coupled with random hexamer primers Phi29 is able to amplify DNA of unknown sequence. This approach allows circular viruses at potentially low concentrations to be amplified without prior knowledge of what is present. RCA can also be utilised to increase the DNA present before concatemer restriction enzyme (RE) digestion, or prior to targeted PCR or next-generation sequencing (NGS). RCA followed by RE digestion have been used to identify the full genomes of episomal BIB (James *et al.*, 2011a).

1.7.5 Paper-based gene sensor

A paper-based gene sensor (PBGS) has recently been developed as a rapid diagnostic test for BBTV, which adapts a lateral flow biosensor to detect BBTV infections after PCR (Wei *et al.*, 2014). Briefly, the PBGS utilises BBTV specific PCR with block oligonucleotides which

are incorporated into the viral DNA. These block oligonucleotides are added to the PBGS and bind to the probes embedded in the paper sensor producing up to two visible lines, the free oligonucleotide line and the PCR viral bound oligonucleotide line, resulting in either a negative (single line) or positive (double line) visual result. This method results in a ten times higher detection using PBGS than electrophoresis following PCR, and is also rapid and low cost (Wei *et al.*, 2014).

1.7.6 Loop-mediated isothermal amplification

The Loop-mediated isothermal amplification (LAMP) method utilises template specific primers and Sybr Green I, resulting in a colour change to quickly identify positive/negative samples (Notomi *et al.*, 2000). This method utilises a strand displacement DNA polymerase and four different primers to rapidly amplify the template DNA without the use of a thermocycler, and with a visual colour change of the reaction (Notomi *et al.*, 2000). LAMP specific primers have been designed and utilised for the detection of both BIB species and BBTV (Peng *et al.*, 2012a; Peng *et al.*, 2012b).

1.7.7 Next-generation sequencing

NGS allows large numbers of sequences from a single sample to be identified and has been utilised in a number of plant samples to detect novel and known viruses (Candresse *et al.*, 2014; Idris *et al.*, 2014; Poojari *et al.*, 2013; Rajeswaran *et al.*, 2014; Seguin *et al.*, 2014). By coupling NGS with RCA, circular sequences are amplified with the likelihood of their detection enhanced. The short sequence reads generated from NGS methods are *de novo* assembled into contigs to identify the potential viral sequences which are present. PCR using back-to-back primers (designed based on the *de novo* assembled contig sequence) can be used to recover circular molecules, which can be cloned and Sanger sequenced to verify the genomes.

1.8 References

Abdel-Salam, S. M., Dahot, M. U. & Sadik, A. S. (2012). Molecular comparative analysis of component 1 (DNA-R) of an Egyptian isolate of banana bunchy top nanovirus isolated from banana aphid (Pentalonia nigronervosa). *Journal of Genetic Engineering and Biotechnology* 10, 55-65.

Abraham, A. D., Varrelmann, M. & Josef Vetten, H. (2012). Three Distinct Nanoviruses, One of Which Represents a New Species, Infect Faba Bean in Ethiopia. *Plant Dis* 96, 1045-1053.

Alberter, B., Ali Rezaian, M. & Jeske, H. (2005). Replicative intermediates of Tomato leaf curl virus and its satellite DNAs. *Virology* 331, 441-448.

Amin, I., Qazi, J., Mansoor, S., Ilyas, M. & Briddon, R. W. (2008). Molecular characterisation of Banana bunchy top virus (BBTV) from Pakistan. *Virus Genes* 36, 191-198.

Anandhi, J., Vijila, C., Viswanath, G. S. & Lokeswari, T. S. (2007). Screening banana plants for banana bunchy top virus with primers specific to Indian isolates. *Journal of Plant Diseases and Protection* 114, 101-107.

Anhalt, M. D. & Almeida, R. P. P. (2008). Effect of temperature, vector life stage, and plant access period on transmission of Banana bunchy top virus to banana. *Phytopathology* 98, 743-748.

Arguello-Astorga, G., Ascencio-Ibáñez, J. T., Dallas, M. B., Orozco, B. M. & Hanley-Bowdoin, L. (2007). High-frequency reversion of geminivirus replication protein mutants during infection. *J Virol* 81, 11005-11015.

Argüello-Astorga, G. R. & Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: Identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 146, 1465-1485.

Aronson, M. N., Meyer, A. D., Györgyey, J., Katul, L., Vetten, H. J., Gronenborn, B. & Timchenko, T. (2000). Clink, a nanovirus-encoded protein, binds both pRB and SKP1. *J Virol* 74, 2967-2972.

Bach, J. & Jeske, H. (2014). Defective DNAs of beet curly top virus from long-term survivor sugar beet plants. *Virus Res* 183, 89-94.

Banerjee, A., Roy, S., Behere, G. T., Roy, S. S., Dutta, S. K. & Ngachan, S. V. (2014). Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. *Virus Res* 183, 41-49.

Bashir, R., Javed, F., Ahmed, R. & Mansoor, S. (2012). Use of rolling circle amplification for the identification of unknown components of banana bunchy top virus from Pakistan. *Pakistan Journal of Life and Social Sciences* 10, 91-97.

Basu, A. N. & Ganguly, B. (1968). A note on the transmission of foorkey disease of large cardamom by the aphid, *Micromyzus kalimpongensis* Basu. *Indian Phytopathology* 21, 127.

Bateson, M. F. & Dale, J. L. (1995). Banana bract mosaic virus: characterisation using potyvirus specific degenerate PCR primers. *Arch Virol* 140, 515-527.

Beetham, P. R., Hafner, G. J., Harding, R. M. & Dale, J. L. (1997). Two mRNAs are transcribed from banana bunchy top virus DNA-1. *J Gen Virol* 78, 229-236.

Beetham, P. R., Harding, R. M. & Dale, J. L. (1999). Banana bunchy top virus DNA-2 to 6 are monocistronic. *Arch Virol* 144, 89-105.

Bell, K. E., Dale, J. L., Ha, C. V., Vu, M. T. & Revill, P. A. (2002). Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Arch Virol* 147, 695-707.

Blackman, R. L. & Eastop, V. P. (2008). *Aphids on the world's herbaceous plants and shrubs*. England: John Wiley and Sons Ltd.

Blanc, S., Drucker, M. & Uzest, M. (2014). Localizing viruses in their insect vectors. In *Annu Rev Phytopathol*, pp. 403-425.

Blomme, G., Jacobsen, K., Ocimati, W., Beed, F., Ntamwira, J., Sivirihauma, C., Ssekiwoko, F., Nakato, V., Kubiriba, J., Tripathi, L., Tinzaara, W., Mbolela, F., Lutete, L. & Karamura, E. (2014). Fine-tuning banana Xanthomonas wilt control options over the past decade in East and Central Africa. *Eur J Plant Pathol* 139, 265-281.

Boevink, P., Chu, P. W. G. & Kees, P. (1995). Sequence of subterranean clover stunt virus DNA: Affinities with the geminiviruses. *Virology* 207, 354-361.

Borth, W., Perez, E., Cheah, K., Chen, Y., Xie, W. S., Gaskill, D., Khalil, S. & Hu, J. S. (2009). Banana bunchy top virus-resistant transgenic banana plants. In *Global perspectives on Asian Challenges*. Guang Zhou.

Borth, W., Perez, E., Cheah, K., Chen, Y., Xie, W. S., Gaskill, D., Khalil, S., Sether, D., Melzer, M., Wang, M., Manshardt, R., Gonsalves, D. & Hu, J. S. (2011). Transgenic banana plants resistant to Banana Bunchy top virus infection. In *Acta Horticulturae*, pp. 449-458.

Briddon, R. W., Bedford, I. D., Tsai, J. H. & Markham, P. G. (1996). Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology* 219, 387-394.

Briddon, R. W., Heydarnejad, J., Khosrowfar, F., Massumi, H., Martin, D. P. & Varsani, A. (2010). Turnip curly top virus, a highly divergent geminivirus infecting turnip in Iran. *Virus Res* 152, 169-175.

Briddon, R. W., Pinner, M. S., Stanley, J. & Markham, P. G. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology* 177, 85-94.

Burns, T. M., Harding R. M. & Dale, J. L. (1995). The genome organisation of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76, 1471-1482.

Burns, T. M., Harding, R. M. & Dale, J. L. (1994). Evidence that banana bunchy top virus has a multiple component genome. *Arch Virol* 137, 371-380.

Candresse, T., Filloux, D., Muhire, B., Julian, C., Galzi, S., Fort, G., Bernardo, P., Daugrois, J. H., Fernandez, E., Martin, D. P., Varsani, A. & Roumagnac, P. (2014). Appearances can be deceptive: Revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS ONE* 9.

Carreel, F., Gonzalez de Leon, D., Lagoda, P., Lanaud, C., Jenny, C., Horry, J. P. & Tezenas du Montcel, H. (2002). Ascertaining maternal and paternal lineage within Musa by chloroplast and mitochondrial DNA RFLP analyses. *Genome* 45, 679-692.

Casado, C. G., Javier Ortiz, G., Padron, E., Bean, S. J., McKenna, R., Agbandje-McKenna, M. & Boulton, M. I. (2004). Isolation and characterization of subgenomic DNAs encapsidated in "single" T = 1 isometric particles of Maize streak virus. *Virology* 323, 164-171.

Chabannes, M., Baurens, F. C., Duroy, P. O., Bocs, S., Vernerey, M. S., Rodier-Goud, M., Barbe, V., Gayral, P. & Iskra-Caruana, M. L. (2013). Three infectious viral species lying in wait in the banana genome. *J Virol* 87, 8624-8637.

Chellappan, S., Kraus, V. B., Kroger, B., Munger, K., Howley, P. M., Phelps, W. C. & Nevins, J. R. (1992). Adeno virus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proc Natl Acad Sci U S A* 89, 4549-4553.

Chen, L. F. & Gilbertson, R. L. (2009). Curtovirus-cucurbit interaction: Acquisition host plays a role in leafhopper transmission in a host-dependent manner. *Phytopathology* 99, 101-108.

Chen, L. I. F., Rojas, M., Kon, T., Gamby, K., Xoconostle Cazares, B. & Golbertson, R. L. (2009). A severe symptom phenotype in tomato in Mali is caused by a reassortant between a novel recombinant begomovirus (Tomato yellow leaf curl Mali virus) and a betasatellite. *Mol Plant Pathol* 10, 415-430.

Chen, Y. F., Chen, J. S., Zhang, H. R., Tang, X. S. & Du, Z. Y. (2007). Molecular evidence and sequence analysis of a natural reassortant between Cucumber mosaic virus subgroup IA and II strains. *Virus Genes* 35, 405-413.

Choudhury, N. R., Malik, P. S., Singh, D. K., Islam, M. N., Kaliappan, K. & Mukherjee, S. K. (2006). The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Res* 34, 6362-6377.

Chu, P. W. G. & Helms, K. (1988). Novel virus-like particles containing circular single-stranded DNAs associated with subterranean clover stunt disease. *Virology* 167, 38-49.

Clérot, D. & Bernardi, F. (2006). DNA helicase activity is associated with the replication initiator protein Rep of tomato yellow leaf curl germinivirus. *J Virol* 80, 11322-11330.

Côte, F. X., Galzi, S., Folliot, M., Lamagnère, Y., Teycheney, P. Y. & Iskra-Caruana, M. L. (2010). Micropropagation by tissue culture triggers differential expression of infectious endogenous

Banana streak virus sequences (eBSV) present in the B genome of natural and synthetic interspecific banana plantains. *Mol Plant Pathol* 11, 137-144.

Creanga, A., Thi Nguyen, D., Gerloff, N., Thi Do, H., Balish, A., Dang Nguyen, H., Jang, Y., Thi Dam, V., Thor, S., Jones, J., Simpson, N., Shu, B., Emery, S., Berman, L., Nguyen, H. T., Bryant, J. E., Lindstrom, S., Klimov, A., Donis, R. O., Davis, C. T. & Nguyen, T. (2013). Emergence of multiple clade 2.3.2.1 influenza A (H5N1) virus subgroups in Vietnam and detection of novel reassortants. *Virology* 444, 12-20.

Cui, X., Tao, X., Xie, Y., Fauquet, C. M. & Zhou, X. (2004). A DNAβ associated with Tomato yellow leaf curl China virus is required for symptom induction. *J Virol* 78, 13966-13974.

D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., Da Silva, C., Jabbari, K., Cardi, C., Poulain, J., Souquet, M., Labadie, K., Jourda, C., Lengellé, J., Rodier-Goud, M., Alberti, A., Bernard, M., Correa, M., Ayyampalayam, S., McKain, M. R., Leebens-Mack, J., Burgess, D., Freeling, M., Mbéguié-A-Mbéguié, D., Chabannes, M., Wicker, T., Panaud, O., Barbosa, J., Hribova, E., Heslop-Harrison, P., Habas, R., Rivallan, R., Francois, P., Poiron, C., Kilian, A., Burthia, D., Jenny, C., Bakry, F., Brown, S., Guignon, V., Kema, G., Dita, M., Waalwijk, C., Joseph, S., Dievart, A., Jaillon, O., Leclercq, J., Argout, X., Lyons, E., Almeida, A., Jeridi, M., Dolezel, J., Roux, N., Risterucci, A. M., Weissenbach, J., Ruiz, M., Glaszmann, J. C., Quétier, F., Yahiaoui, N. & Wincker, P. (2012). The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature* 488, 213-217.

D'Hont, A., Paget-Goy, A., Escoute, J. & Carreel, F. (2000). The interspecific genome structure of cultivated banana, Musa spp. revealed by genomic DNA in situ hybridization. *Theor Appl Genet* 100, 177-183.

Dahal, G., Hughes, J. D., Thottappilly, G. & Lockhart, B. E. L. (1998). Effect of temperature on symptom expression and reliability of banana streak badnavirus detection in naturally infected plantain and banana (Musa spp.). *Plant Dis* 82, 16-21.

Dale, J. L. (1987). Banana Bunchy Top - An economically important tropical plant-virus disease. *Adv Virus Res* 33, 301-325.

Daniells, J. W., Geering, A. D. W., Bryde, N. J. & Thomas, J. E. (2001). The effect of Banana streak virus on the growth and yield of dessert bananas in tropical Australia. *Ann Appl Biol* 139, 51-60.

Davey, M. W., Gudimella, R., Harikrishna, J. A., Sin, L. W., Khalid, N. & Keulemans, J. (2013). A draft Musa balbisiana genome sequence for molecular genetics in polyploid, inter- and intra-specific Musa hybrids. *BMC Genomics* 14.

Davis, R. I., Kokoa, P., Jones, L. M., MacKie, J., Constable, F. E., Rodoni, B. C., Gunua, T. G. & Rossel, J. B. (2012). A new wilt disease of banana plants associated with phytoplasmas in Papua New Guinea (PNG). *Australasian Plant Disease Notes* 7, 91-97.

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, M., Abdoul-Karime, A. L., Abdou-Chakour, C., Reynaud, B., Harkins, G. W., Varsani, A., Martin, D. P. & Lett, J. M. (2012). East African cassava mosaic-like viruses from Africa to Indian ocean islands: Molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC Evol Biol*, 228.

Dekker, E. L., Woolston, C. J., Xue, Y., Cox, B. & Mullineaux, P. M. (1991). Transcript mapping reveals different expression strategies for the bicistronic RNAs of the geminivirus wheat dwarf virus. *Nucleic Acids Res* 19, 4075-4081.

Denham, T., Haberle, S. & Lentfer, C. (2004). New evidence and revised interpretations of early agriculture in Highland New Guinea. *Antiquity* 78, 839-857.

Dheepa, R. & Paranjothi, S. (2010). Transmission of Cucumber Mosaic Virus (CMV) infecting banana by aphid and mechanical methods. *Emirates Journal of Food and Agriculture* 22, 117-129.

Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus. *J Virol* 82, 957-965.

Duffy, S. & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* 90, 1539-1547.

Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9, 267-276.

Dugdale, B., Beetham, P. R., Becker, D. K., Harding, R. M. & Dale, J. L. (1998). Promoter activity associated with the intergenic regions of banana bunchy top virus DNA-1 to -6 in transgenic tobacco and banana cells. *J Gen Virol* 79, 2301-2311.

Dugdale, B., Mortimer, C. L., Kato, M., James, T. A., Harding, R. M. & Dale, J. L. (2013). In plant activation: An inducible, hyperexpression platformfor recombinant protein production in plants. *Plant Cell* 25, 2429-2443.

Elayabalan, S., Kalaiponmani, K., Subramaniam, S., Selvarajan, R., Panchanathan, R., Muthuvelayoutham, R., Kumar, K. K. & Balasubramanian, P. (2013). Development of Agrobacterium-mediated transformation of highly valued hill banana cultivar Virupakshi (AAB) for resistance to BBTV disease. *World Journal of Microbiology and Biotechnology* 29, 589-596.

Elayabalan, S., Subramaniam, S. & Selvarajan, R. (2015). Banana bunchy top disease (BBTD) symptom expression in banana and strategies for transgenic resistance: A review. *Emirates Journal of Food and Agriculture* 27, 55-74.

Fahmy, T. (1927). *Minerals and Agriculture in Egypt Bulletin*, 30.

FAOSTAT (2012). Food and Agriculture Organisation of the United Nations Statistics Division. 2012.

Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M. & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Arch Virol* 153, 783-821.

Firth, C., Charleston, M. A., Duffy, S., Shapiro, B. & Holmes, E. C. (2009). Insights into the evolutionary history of an emerging livestock pathogen: Porcine circovirus 2. *J Virol* 83, 12813-12821.

Fischer, A., Strohmeier, S., Krenz, B. & Jeske, H. (2015). Evolutionary liberties of the Abutilon mosaic virus cluster. *Virus Genes* 50, 63-70.

Foottit, R., Maw, H., Pike, K. & Miller, R. (2010). The identity of Pentalonia nigronervosa Coquerel and P. caladii van der Goot (Hemiptera: Aphididae) based on molecular and morphometric analysis. *Zootaxa* 2358, 25-38.

Franz, A. W. E., van der Wilk, F., Verbeek, M., Dullemans, A. M. & van den Heuvel, J. F. (1999). Faba bean necrotic yellows virus (genus Nanovirus) requires a helper factor for its aphid transmission. *Virology* 262, 210-219.

Frischmuth, T. & Stanley, J. (1992). Characterization of beet curly top virus subgenomic DNA localizes sequences required for replication. *Virology* 189, 808-811.

Fu, H. C., Hu, J. M., Hung, T. H., Su, H. J. & Yeh, H. H. (2009). Unusual events involved in banana bunchy top virus strain evolution. *Phytopathology* 99, 812-822.

Furuya, N., Dizon, T. O. & Natsuaki, K. T. (2006). Molecular Characterization of Banana bunchy top virus and Cucumber mosaic virus from Abaca (Musa textilis Nee). *Journal of Agricultural Science-Tokyo University of Agriculture* 51, 92-101.

Furuya, N., Kawano, S. & Natsuaki, K. T. (2005). Characterization and genetic status of banana bunchy top virus isolated from Okinawa, Japan. *J Gen Plant Pathol* 71, 68-73.

Furuya, N., Somowiyarjo, S. & Natsuaki, K. (2004). Virus detection from local banana cultivars and the first molecular characterization of Banana bunchy top virus in Indonesia. *Journal of Agricultural Science-Tokyo University of Agriculture* 49, 75-81.

García-Bastidas, F., Ordóñez, N., Konkol, J., Al-Qasim, M., Naser, Z., Abdelwali, M., Salem, N., Waalwijk, C., Ploetz, R. C. & Kema, G. H. J. (2014). First report of Fusarium oxysporum f. sp. cubense tropical race 4 associated with panama disease of banana outside Southeast Asia. *Plant Dis* 98, 694.

Gayral, P., Noa-Carrazana, J. C., Lescot, M., Lheureux, F., Lockhart, B. E. L., Matsumoto, T., Piffanelli, P. & Iskra-Caruana, M. L. (2008). A single Banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. *J Virol* 82, 6697-6710.

Geering, A. D. W., Maumus, F., Copetti, D., Choisne, N., Zwickl, D. J., Zytnicki, M., McTaggart, A. R., Scalabrin, S., Vezzulli, S., Wing, R. A., Quesneville, H. & Teycheney, P.-Y. (2014).

Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nat Commun* 5.

Geering, A. D. W., McMichael, L. A., Dietzgen, R. G. & Thomas, J. E. (2000). Genetic diversity among banana streak virus isolates from Australia. *Phytopathology* 90, 921-927.

Geering, A. D. W., Olszewski, N. E., Dahal, G., Thomas, J. E. & Lockhart, B. E. L. (2001). Analysis of the distribution and structure of integrated banana streak virus DNA in a range of Musa cultivars. *Mol Plant Pathol* 2, 207-213.

Geering, A. D. W., Olszewski, N. E., Harper, G., Lockhart, B. E. L., Hull, R. & Thomas, J. E. (2005a). Banana contains a diverse array of endogenous badnaviruses. *J Gen Virol* 86, 511-520.

Geering, A. D. W., Parry, J. N. & Thomas, J. E. (2011). Complete genome sequence of a novel badnavirus, banana streak IM virus. *Arch Virol* 156, 733-737.

Geering, A. D. W., Pooggin, M. M., Olszewski, N. E., Lockhart, B. E. L. & Thomas, J. E. (2005b). Characterisation of banana streak mysore virus and evidence that its DNA is integrated in the B genome of cultivated Musa. *Arch Virol* 150, 787-796.

Geering, A. D. W. & Thomas, J. E. (1997). Search for alternative hosts of banana bunchy top virus in Australia. *Australas Plant Pathol* 26, 250-254.

Gharouni Kardani, S., Heydarnejad, J., Zakiaghl, M., Mehrvar, M., Kraberger, S. & Varsani, A. (2013). Diversity of Beet curly top Iran virus isolated from different hosts in Iran. *Virus Genes* 46, 571-575.

Gorbalenya, A. E., Koonin, E. V. & Wolf, Y. I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett* 262, 145-148.

Grigoras, I., Ginzo, A. I. d. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.-J. & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J Gen Virol* 95, 1178-1191.

Grigoras, I., Gronenborn, B. & Vetten, H. J. (2010a). First Report of a Nanovirus Disease of Pea in Germany. *Plant Dis* 94, 642-642.

Grigoras, I., Timchenko, T., Grande-Perez, A., Katul, L., Vetten, H. J. & Gronenborn, B. (2010b). High variability and rapid evolution of a nanovirus. *J Virol* 84, 9105-9117.

Grigoras, I., Timchenko, T., Katul, L., Grande-Perez, A., Vetten, H. J. & Gronenborn, B. (2009). Reconstitution of authentic nanovirus from multiple cloned DNAs. *J Virol* 83, 10778-10787.

Gronenborn, B. (2004). Nanoviruses: Genome organisation and protein function. *Vet Microbiol* 98, 103-109.

Gu, H., Zhang, C. & Ghabrial, S. A. (2007). Novel naturally occurring Bean pod mottle virus reassortants with mixed heterologous RNA1 genomes. *Phytopathology* 97, 79-86.

Guan, Y., Peiris, J. S. M., Lipatov, A. S., Ellis, T. M., Dyrting, K. C., Krauss, S., Zhang, L. J., Webster, R. G. & Shortridge, K. F. (2002). Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. *Proc Natl Acad Sci U S A* 99, 8950-8955.

Guo, W., Jiang, T., Zhang, X., Li, G. & Zhou, X. (2008). Molecular variation of satellite DNAß molecules associated with Malvastrum yellow vein virus and their role in pathogenicity. *Appl Environ Microbiol* 74, 1909-1913.

Gutierrez, C. (1999). Geminivirus DNA replication. *Cellular and Molecular Life Sciences* 56, 313-329.

Hadfield, J., Thomas, J. E., Schwinghamer, M. W., Kraberger, S., Stainton, D., Dayaram, A., Parry, J. N., Pande, D., Martin, D. P. & Varsani, A. (2012). Molecular characterisation of dicot-infecting mastreviruses from Australia. *Virus Res* 166, 13-22.

Hafner, G. J., Harding, R. M. & Dale, J. L. (1995). Movement and transmission of banana bunchy top virus DNA component one in bananas. *J Gen Virol* 76, 2279-2285.

Hafner, G. J., Harding, R. M. & Dale, J. L. (1997a). A DNA primer associated with banana bunchy top virus. *J Gen Virol* 78, 479-486.

Hafner, G. J., Stafford, M. R., Wolter, L. C., Harding, R. M. & Dale, J. L. (1997b). Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J Gen Virol* 78, 1795-1799.

Häkkinen, M. (2013). Reappraisal of sectional taxonomy in Musa (Musaceae). *Taxon* 62, 809-813.

Harding, R. M., Burns, T. M. & Dale, J. L. (1991). Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *J Gen Virol* 72, 225-230.

Harding, R. M., Burns, T. M., Hafner, G., Dietzgen, R. G. & Dale, J. L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *J Gen Virol* 74, 323-328.

Harkins, G. W., Martin, D. P., Christoffels, A. & Varsani, A. (2014). Towards inferring the global movement of beak and feather disease virus. *Virology* 450-451, 24-33.

Harkins, G. W., Martin, D. P., Duffy, S., Monjane, A. L., Shepherd, D. N., Windram, O. P., Owor, B. E., Donaldson, L., van Antwerpen, T., Sayed, R. A., Flett, B., Ramusi, M., Rybicki, E. P., Peterschmitt, M. & Varsani, A. (2009). Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. *J Gen Virol* 90, 3066-3074.

Harper, G. & Hull, R. (1998). Cloning and sequence analysis of banana streak virus DNA. *Virus Genes* 17, 271-278.

Harper, G., Osuji, J. O., Heslop-Harrison, J. S. & Hull, R. (1999). Integration of banana streak badnavirus into the Musa genome: Molecular and cytogenetic evidence. *Virology* 255, 207-213.

He, Z. F., Li, H. P., Xiao, H. G. & Fan, H. Z. (2000). Cloning and sequencing of DNA component 1 of two BBTV strains. *Chih Wu Ping Li Hsueh Pao* 30, 364-369.

He, Z. F., Li, H. P., Xiao, H. G. & Fan, H. Z. (2001a). Cloning and sequencing of DNA component 3 of two BBTV strains. *Nongye Shengwu Jisu Xuebao* 9, 145-148.

He, Z. F., Li, H. P., Xiao, H. G. & Fan, H. Z. (2001b). Cloning and sequencing of DNA component 6 of two BBTV strains. *Hua Nan Nung Yeh Ta Hsueh Hsueh Pao* 22, 33-36.

Hernández-Zepeda, C., Varsani, A. & Brown, J. K. (2013). Intergeneric recombination between a new, spinach-infecting curtovirus and a new geminivirus belonging to the genus Becurtovirus: First New World exemplar. *Arch Virol* 158, 2245-2254.

Herrera-Valencia, V. A., Dugdale, B., Harding, R. M. & Dale, J. L. (2006). An iterated sequence in the genome of banana bunchy top virus is essential for efficient replication. *J Gen Virol* 87, 3409-3412.

Heydarnejad, J., Keyvani, N., Razavinejad, S., Massumi, H. & Varsani, A. (2013). Fulfilling Koch's postulates for beet curly top Iran virus and proposal for consideration of new genus in the family Geminiviridae. *Arch Virol* 158, 435-443.

Hipp, K., Rau, P., Schäfer, B., Gronenborn, B. & Jeske, H. (2014). The RXL motif of the African cassava mosaic virus Rep protein is necessary for rereplication of yeast DNA and viral infection in plants. *Virology* 462-463, 189-198.

Hohn, T., Richert-Pöggeler, K., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C., Teycheney, P.-Y., Iskra-Caruana, M.-L. & Hull, R. (2008). Evolution of Integrated Plant Viruses. In *Plant Virus Evolution*, pp. 53-81. Edited by M. Roossinck: Springer Berlin Heidelberg.

Höhnle, M., Höfer, P., Bedford, I. D., Briddon, R. W., Markham, P. G. & Frischmuth, T. (2001). Exchange of three amino acids in the coat protein results in efficient whitefly transmission of a nontransmissible Abutilon mosaic virus isolate. *Virology* 290, 164-171.

Hooks, C. R. R., Wright, M. G., Kabasawa, D. S., Manandhar, R. & Almeida, R. P. P. (2008). Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. *Annuals of Applied Biology* 153, 1-9.

Horn, J., Lauster, S., Krenz, B., Kraus, J., Frischmuth, T. & Jeske, H. (2011). Ambivalent effects of defective DNA in beet curly top virus-infected transgenic sugarbeet plants. *Virus Res* 158, 169-178.

Horser, C. L., Harding, R. M. & Dale, J. L. (2001a). Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *J Gen Virol* 82, 459-464.

Horser, C. L., Karan, M., Harding, R. M. & Dale, J. L. (2001b). Additional Rep-encoding DNAs associated with banana bunchy top virus. *Arch Virol* 146, 71-86.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. (2007). Reassortment and concerted evolution in Banana bunchy top virus genomes. *J Virol* 81, 1746-1761.

Hu, J. S., Wang, M., Sether, D., Xie, W. & Leonhardt, K. W. (1996). Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (Pentalonia nigronervosa). *Ann Appl Biol* 128, 55-64.

Hughes, A. L. (2004). Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. *Mol Phylogenet Evol* 30, 287-294.

Hyder, M. Z. (2009). Sequencing and Genetic Characterization of major components of Banana Bunchy Top Virus. In *Department of Biochemistry*. Pakistan: Arid Agriculture University.

Hyder, M. Z., Raza, S. Q., Hameed, S., Khalid, S. & Saqlan Naqvi, S. M. (2007). Phylogenetic relationship of TJ1 isolate of Banana bunchy top virus from Pakistan by DNA-R sequence analysis. *Can J Plant Pathol* 29, 63-68.

Hyder, M. Z., Shah, S. H., Hameed, S. & Naqvi, S. M. S. (2011). Evidence of recombination in the Banana bunchy top virus genome. *Infect Genet Evol* 11, 1293-1300.

Idris, A., Al-Saleh, M., Piatek, M. J., Al-Shahwan, I., Ali, S. & Brown, J. K. (2014). Viral metagenomics: Analysis of begomoviruses by illumina high-throughput sequencing. *Viruses* 6, 1219-1236.

Idris, A. M. & Brown, J. K. (2004). Cotton leaf crumple virus is a distinct western hemisphere begomovirus species with complex evolutionary relationships indicative of recombination and reassortment. *Phytopathology* 94, 1068-1074.

Idris, A. M., Shahid, M. S., Briddon, R. W., Khan, A. J., Zhu, J. K. & Brown, J. K. (2011). An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. *J Gen Virol* 92, 706-717.

Irish, B. M., Goenaga, R., Rios, C., Chavarria-Carvajal, J. & Ploetz, R. (2013). Evaluation of banana hybrids for tolerance to black leaf streak (Mycosphaerella fijiensis Morelet) in Puerto Rico. *Crop Prot* 54, 229-238.

Iskra-Caruana, M. L., Baurens, F. C., Gayral, P. & Chabannes, M. (2010). A four-partner plant-virus interaction: Enemies can also come from within. *Mol Plant-Microbe Interact* 23, 1394-1402.

Islam, M. N., Naqvi, A. R., Jan, A. T. & Haq, Q. M. R. (2010). Genetic Diversity and Possible Evidence of Recombination among Banana Bunchy Top Virus (BBTV) Isolates. *International Research Journal of Microbiology* Vol 1, 001-012.

James, A. P. (2011). Viruses of Banana in East Africa: Queensland University of Technology.

James, A. P., Geijskes, R. J., Dale, J. L. & Harding, R. M. (2011a). Development of a novel rolling-circle amplification technique to detect Banana streak virus that also discriminates between integrated and episomal virus sequences. *Plant Dis* 95, 57-62.

James, A. P., Geijskes, R. J., Dale, J. L. & Harding, R. M. (2011b). Molecular characterisation of six badnavirus species associated with leaf streak disease of banana in East Africa. *Ann Appl Biol* 158, 346-353.

Jarret, R. L., Gawel, N., Whittemore, A. & Sharrock, S. (1992). RFLP-based phylogeny of Musa species in Papua New Guinea. *Theor Appl Genet* 84, 579-584.

Jeske, H., Kober, S., Schäfer, B. & Strohmeier, S. (2014). Circomics of Cuban geminiviruses reveals the first alpha-satellite DNA in the Caribbean. *Virus Genes* 49, 312-324.

Jeske, H., Lütgemeier, M. & Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO Journal* 20, 6158-6167.

Jovel, J., Preiss, W. & Jeske, H. (2007). Characterization of DNA intermediates of an arising geminivirus. *Virus Res* 130, 63-70.

Julian, L., Piasecki, T., Chrzastek, K., Walters, M., Muhire, B., Harkins, G. W., Martin, D. P. & Varsani, A. (2013). Extensive recombination detected among beak and feather disease virus isolates from breeding facilities in Poland. *J Gen Virol* 94, 1086-1095.

Jun, T. E. Z. & Zhi-Xin, L. (2005). Cloning and sequencing of DNA components of Banana bunchy top virus Hainan isolate. *Chinese Journal of Agricultural Biotechnology* 2, 91-97.

Karan, M., Harding, R. M. & Dale, J. L. (1994). Evidence for two groups of banana bunchy top virus isolates. *J Gen Virol* 75, 3541-3546.

Karan, M., Harding, R. M., Dale, J. L. (1997). Association of Banana Bunchy Top Virus DNA Components 2 to 6 with Bunchy Top Disease. *Molecular Plant Pathology On-line*.

Katul, L., Timchenko, T., Gronenborn, B. & Vetten, H. J. (1998). Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *J Gen Virol* 79, 3101-3109.

King, A. M. Q., Lefkowitz, E., Adams, M. J. & Carstens, E. B. (2011). Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses. Burlington: Elsevier Science.

Kraberger, S., Harkins, G. W., Kumari, S. G., Thomas, J. E., Schwinghamer, M. W., Sharman, M., Collings, D. A., Briddon, R. W., Martin, D. P. & Varsani, A. (2013). Evidence that dicot-infecting mastreviruses are particularly prone to inter-species recombination and have likely been circulating in Australia for longer than in Africa and the Middle East. *Virology* 444, 282-291.

Kraberger, S., Kumari, S. G., Hamed, A. A., Gronenborn, B., Thomas, J. E., Sharman, M., Harkins, G. W., Muhire, B. M., Martin, D. P. & Varsani, A. (2015). Molecular diversity of Chickpea chlorotic dwarf virus in Sudan: High rates of intra-species recombination - a driving force in the emergence of new strains. *Infection, Genetics and Evolution* 29, 203-215.

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I. & Simon, R. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1-7.

Kumar, J., Kumar, J., Singh, S. P. & Tuli, R. (2014). Association of satellites with a mastrevirus in natural infection: Complexity of Wheat dwarf India virus disease. *J Virol* 88, 7093-7104.

Kumar, P., Hanna, R., Alabi, O., Soko, M., Oben, T., Vangu, G. & Naidu, R. (2011). Banana bunchy top virus in sub-Saharan Africa: investigations on virus distribution and diversity. *Virus Res* 159, 171.

Kusumoto, S., Aeny, T. N., Mujimu, S., Ginting, C., Tsuge, T., Tsuyumu, S. & Takikawa, Y. (2004). Occurrence of blood disease of banana in Sumatra, Indonesia. *J Gen Plant Pathol* 70, 45-49.

La, P., Cai, W. Q. & Fang, R. X. (2000). Cloning and sequencing of banana bunchy top virus genome component I. *Chinese Journal of Virology* 16, 158-161.

Lageix, S., Catrice, O., Deragon, J. M., Gronenborn, B., Pélissier, T. & Ramirez, B. C. (2007). The nanovirus-encoded clink protein affects plant cell cycle regulation through interaction with the retinoblastoma-related protein. *J Virol* 81, 4177-4185.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J. & Gronenborn, B. (1995). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc Natl Acad Sci U S A* 92, 3879-3883.

Lava Kumar, P., Selvarajan, R., Iskra-Caruana, M. L., Chabannes, M. & Hanna, R. (2014). Biology, Etiology, and Control of Virus Diseases of Banana and Plantain. In *Adv Virus Res*.

Le Provost, G., Iskra-Caruana, M.-L., Acina, I. & Teycheney, P.-Y. (2006). Improved detection of episomal Banana streak viruses by multiplex immunocapture PCR. *J Virol Methods* 137, 7-13.

Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009). Widely Conserved Recombination Patterns among Single-Stranded DNA Viruses. *J Virol* 83, 2697-2707.

Lheureux, F., Laboureau, N., Muller, E., Lockhart, B. E. L. & Iskra-Caruana, M. L. (2007). Molecular characterization of banana streak acuminata Vietnam virus isolated from Musa acuminata siamea (banana cultivar). *Arch Virol* 152, 1409-1416.

Li, Z., Xie, Y. & Zhou, X. (2005). Tobacco curly shoot virus DNAβ is not necessary for infection but intensifies symptoms in a host-dependent manner. *Phytopathology* 95, 902-908.

Lockhart, B. (1990). Evidence for a Double-Stranded Circular DNA Genome in a Second Group of Plant Viruses. *Phytopathology* 80, 127-131.

Lockhart, B. E., Menke, J., Dahal, G. & Olszewski, N. E. (2000). Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *J Gen Virol* 81, 1579-1585.

Lockhart, B. E. L. (1968). Purification and serology of a bacilliform virus associated with banana streak disease. *Phytopathology* 76, 995-999.

Londoño, A., Riego-Ruiz, L. & Argüello-Astorga, G. R. (2010). DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Arch Virol* 155, 1033-1046.

Lozano-Durán, R., Rosas-Díaz, T., Gusmaroli, G., Luna, A. P., Taconnat, L., Deng, X. W. & Bejarano, E. R. (2011). Geminiviruses subvert ubiquitination by altering CSN-mediated

derubylation of SCF E3 ligase complexes and inhibit jasmonate signaling in Arabidopsis thaliana. *Plant Cell* 23, 1014-1032.

Magee, C. J. P. (1927). Investigation on the bunchy top disease of the banana. *Bulletin of the Council for Scientific and Industrial Research in Australia* 30.

Magee, C. J. P. (1940). Transmission studies on the Banana Bunchy-top Virus. *Journal of the Australian Institute of Agricultural Science* 6, 109-110.

Magnaye, L. V. & Espino, R. R. C. (1990). Note: Banana bract mosaic, a new disease of banana. *The Philippine Agriculturist* 73, 55-59.

Mandal, B., Mandal, S., Pun, K. B. & Varma, A. (2004). First report of the association of a nanovirus with foorkey disease of large cardamom in India. *Plant Dis* 88, 428-428.

Mandal, B., Shilpi, S., Barman, A. R., Mandal, S. & Varma, A. (2013). Nine novel DNA components associated with the foorkey disease of large cardamom: Evidence of a distinct babuvirus species in Nanoviridae. *Virus Res* 178, 297-305.

Manickam, K., Doraiswamy, S., Ganapathy, T. & Rabindran, R. (2002). Early detection of banana bunchy top virus in India using polymerase chain reaction. *Acta Phytopathol Entomol Hung* 37, 9-16.

Manoranjitham, S., Kavino, M., Thiribhuvanamala, G., Ganapathy, T., Rabindran, R. & Kumar, N. (2012). Detection of banana streak virus (BSV) Tamil Nadu isolate (India) and its serological relationship with other badna viruses. *Afr J Biotechnol* 11, 14632-14637.

Maoka, T., Hayano, Y. S., Iwasaki, M., Yoshida, K. & Masuta, C. (2010). Mixed infection in tomato to ensure frequent generation of a natural reassortant between two subgroups of Cucumber mosaic virus. *Virus Genes* 40, 148-150.

Martin, D. P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P. & Varsani, A. (2011). Recombination in eukaryotic single stranded DNA viruses. *Viruses* 3, 1699-1738.

Matthews, R. E. F. (1981). The Classification and Nomenclature of Viruses: Summary of Results of Meetings of the International Committee on Taxonomy of Viruses in Strasbourg, August 1981. *Intervirology* 16, 53-60.

Mbida Mindzie, C., Doutrelepont, H., Vrydaghs, L., Swennen, R. L., Swennen, R. J., Beeckman, H., de Langhe, E. & de Maret, P. (2001). First archaeological evidence of banana cultivation in central Africa during the third millennium before present. *Veg Hist Archaeobot* 10, 1-6.

Meyer, J. B., Kasdorf, G. G. F., Nel, L. H. & Pietersen, G. (2008). Transmission of activated-episomal Banana streak OL (badna)virus (BSOLV) to cv. Williams banana (Musa sp.) by three mealybug species. *Plant Dis* 92, 1158-1163.

Monjane, A. L., Martin, D. P., Lakay, F., Muhire, B. M., Pande, D., Varsani, A., Harkins, G., Shepherd, D. N. & Rybicki, E. P. (2014). Extensive recombination-induced disruption of genetic interactions is highly deleterious but can be partially reversed by small numbers of secondary recombination events. *J Virol* 88, 7843-7851.

Monjane, A. L., Pande, D., Lakay, F., Shepherd, D. N., Van Der Walt, E., Lefeuvre, P., Lett, J. M., Varsani, A., Rybicki, E. P. & Martin, D. P. (2012). Adaptive evolution by recombination is not associated with increased mutation rates in Maize streak virus. *BMC Evol Biol* 12.

Mouliom-Pefoura, A., Lassoudière, A., Foko, J. & Fontem, D. A. (1996). Comparison of development of Mycosphaerella fijiensis and Mycosphaerella musicola on banana and plantain in the various ecological zones in Cameroon. *Plant Dis* 80, 950-954.

Muhire, B., Martin, D., Brown, J., Navas-Castillo, J., Moriones, E., Zerbini, F. M., Rivera-Bustamante, R., Malathi, V. G., Briddon, R. & Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). *Arch Virol* 158, 1411-1424.

Muhire, B. M., Golden, M., Murrell, B., Lefeuvre, P., Lett, J. M., Gray, A., Poon, A. Y. F., Ngandu, N. K., Semegni, Y., Tanov, E. P., Monjane, A. L., Harkins, G. W., Varsani, A., Shepherd, D. N. & Martin, D. P. (2014a). Evidence of pervasive biologically functional secondary structures within the genomes of eukaryotic single-stranded DNA viruses. *J Virol* 88, 1972-1989.

Muhire, B. M., Varsani, A. & Martin, D. P. (2014b). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* 9, e108277.

Mullineaux, P. M., Guerineau, F. & Accotto, G. P. (1990). Processing of complementary sense RNAs of Digitaria streak virus in its host and in transgenic tobacco. *Nucleic Acids Res* 18, 7259-7265.

Mwangi, M. & Nakato, V. (2009). Key factors responsible for the xanthomonas wilt epidemic on banana in East and Central Africa. In *Acta Horticulturae*, pp. 395-404.

Nawaz-ul-Rehman, M. S., Nahid, N., Mansoor, S., Briddon, R. W. & Fauquet, C. M. (2010). Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* 405, 300-308.

Ndowora, T., Dahal, G., Lafleur, D., Harper, G., Hull, R., Olszewski, N. E. & Lockhart, B. (1999). Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. *Virology* 255, 214-220.

Ndungo, V., Eden-Green, S., Blomme, G., Crozier, J. & Smith, J. J. (2006). Presence of banana xanthomonas wilt (Xanthomonas campestris pv. musacearum) in the Democratic Republic of Congo (DRC). *Plant Pathology* 55, 294.

Nelson, M. I., Viboud, C., Simonsen, L., Bennett, R. T., Griesemer, S. B., St. George, K., Taylor, J., Spiro, D. J., Sengamalay, N. A., Ghedin, E., Taubenberger, J. K. & Holmes, E. C. (2008). Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathogens* 4.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N. & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res* 28.

Nwakanma, D. C., Pillay, M., Okoli, B. E. & Tenkouano, A. (2003). PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in Musa L. *Theor Appl Genet* 108, 154-159.

Orozco, B. M. & Hanley-Bowdoin, L. (1998). Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *J Biol Chem* 273, 24448-24456.

Owor, B. E., Martin, D. P., Shepherd, D. N., Edema, R., Monjane, A. L., Rybicki, E. P., Thomson, J. A. & Varsani, A. (2007a). Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. *J Gen Virol* 88, 3154-3165.

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., Martin, D. P. & Varsani, A. (2007b). Successful application of FTA® Classic Card technology and use of bacteriophage phi 29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *J Virol Methods* 140, 100-105.

Paprotka, T., Boiteux, L. S., Fonseca, M. E. N., Resende, R. O., Jeske, H., Faria, J. C. & Ribeiro, S. G. (2010a). Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Res* 149, 224-233.

Paprotka, T., Metzler, V. & Jeske, H. (2010b). The first DNA 1-like α satellites in association with New World begomoviruses in natural infections. *Virology* 404, 148-157.

Parmenter, K. S. & Thomas, J. E. (2014). Persistence of infectivity in herbicide/aphicide-injected *Banana bunchy top virus* infected plants. In *29th International Horticulture Congress 2014*. Brisbane, Australia.

Patil, B. L. & Dasgupta, I. (2006). Defective interfering DNAs of plant viruses. *Crit Rev Plant Sci* 25, 47-64.

Peeters, N., Guidot, A., Vailleau, F. & Valls, M. (2013). Ralstonia solanacearum, a widespread bacterial plant pathogen in the post-genomic era. *Mol Plant Pathol* 14, 651-662.

Pekin, J. (2006). Queensland Government Position Paper: Repeal of the *Banana Industry Protection Act 1989*. Edited by D. o. P. I. a. F. The State of Queensland.

Peng, J., Fan, Z. & Huang, J. (2012a). Rapid Detection of Banana Streak Virus by Loop-mediated Isothermal Amplification Assay in South China. *Journal of Phytopathology* 160, 248-250.

Peng, J., Zhang, J., Xia, Z., Li, Y., Huang, J. & Fan, Z. (2012b). Rapid and sensitive detection of Banana bunchy top virus by loop-mediated isothermal amplification. *J Virol Methods* 185, 254-258.

Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J. P., Jenny, C., Lebot, V., Risterucci, A. M., Tomekpe, K., Doutrelepont, H., Ball,

T., Manwaring, J., de Maret, P. & Denham, T. (2011). Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc Natl Acad Sci U S A* 108, 11311-11318.
Pillay, M., Nwakanma, D. C. & Tenkouano, A. (2000). Identification of RAPD markers linked to A and B genome sequences in Musa L. *Genome* 43, 763-767.
Pillay, M., Ogundiwin, E., Tenkouano, A. & Dolezel, J. (2006). Ploidy and genome composition of Musa germplasm at the International Institute of Tropical Agriculture (IITA). *Afr J Biotechnol* 5, 1224-1232.
Pinili, M. S., Nagashima, I., Dizon, T. O. & T. Natsuaki, K. (2013). Cross-Transmission and New Alternate Hosts of *Banana bunchy top virus*. *Tropical Agriculture and Development* 57, 1-7.
Pinili, M. S., Nyana, D., Suastika, G. & Natsuaki, K. (2011). Molecular analysis of Banana bunchy top virus first isolated in Bali, Indonesia. *J Agric Sci Tokyo Univ Agric* 56, 125-134.
Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* 82, 655-665.
PlantHealthAustralia (2006). Bacterial Wilt of Banana diagnostics manual. Australia: Plant Health Australia.
Ploetz, R. C. (2006). Fusarium wilt of banana is caused by several pathogens referred to as Fusarium oxysporum f. sp. cubense. *Phytopathology* 96, 653-656.
Poojari, S., Alabi, O. J., Fofanov, V. Y. & Naidu, R. A. (2013). A Leafhopper-Transmissible DNA Virus with Novel Evolutionary Lineage in the Family Geminiviridae Implicated in Grapevine Redleaf Disease by Next-Generation Sequencing. *PLoS ONE* 8.
Preiss, W. & Jeske, H. (2003). Multitasking in replication is common among geminiviruses. *J Virol* 77, 2972-2980.
ProMusa (2014). Banana cultivar checklist: Musapedia, the banana knowledge compendium.
Qiu, W. P. & Moyer, J. W. (1999). Tomato spotted wilt tospovirus adapts to the TSWV N gene-derived resistance by genome reassortment. *Phytopathology* 89, 575-582.
Quito-Avila, D. F., Ibarra, M. A., Alvarez, R. A., Ratti, M. F., Espinoza, L., Cevallos-Cevallos, J. M. & Peralta, E. L. (2013). First report of banana bract mosaic virus in 'Cavendish' banana in Ecuador. *Plant Dis* 97, 1003.
Rajeswaran, R., Seguin, J., Chabannes, M., Duroy, P. O., Laboureau, N., Farinelli, L., Iskra-Caruana, M. L. & Pooggin, M. M. (2014). Evasion of short interfering RNA-directed antiviral silencing in Musa acuminata persistently infected with six distinct banana streak pararetroviruses. *J Virol* 88, 11516-11528.
Ram, R. D. & Summanwar, A. S. (1984). Colocasia esculenta (L.) schott. A reservoir of bunchy top disease of banana. *Curr Sci* 53, 145-146.
Randles, J., Julia, J., Calvez, C. & Dollet, M. (1986). Association of single-stranded DNA with the foliar decay disease of coconut palm in Vanuatu. *Phytopathology* 76, 889-894.
Razavinejad, S. & Heydarnejad, J. (2013). Transmission and natural hosts of turnip curly top virus. *Iran J Plant Pathol* 49, 27-28.
Razavinejad, S., Heydarnejad, J., Kamali, M., Massumi, H., Kraberger, S. & Varsani, A. (2013). Genetic diversity and host range studies of turnip curly top virus. *Virus Genes* 46, 345-353.
Reeder, R. H., Muhinyuza, J. B., Opolot, O., Aritua, V., Crozier, J. & Smith, J. (2007). Presence of banana bacterial wilt (Xanthomonas campestris pv. musacearum) in Rwanda. *Plant Pathology* 56, 1038.
Rez, R. P., Calleros, L., Marandino, A., Sarute, N., Iraola, G., Grecco, S., Blanc, H., Vignuzzi, M., Isakov, O., Shomron, N., Carrau, L., Hernández, M., Francia, L., Sosa, K., Tomás, G. & Panzera, Y. (2014). Phylogenetic and genome-wide deep-sequencing analyses of Canine parvovirus reveal co-infection with field variants and emergence of a recent recombinant strain. *PLoS ONE* 9.
Richert-Pöggeler, K. R., Noreen, F., Schwarzacher, T., Harper, G. & Hohn, T. (2003). Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. *EMBO Journal* 22, 4836-4845.
Rohde, W., Randles, J. W., Langridge, P. & Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* 176, 648-651.

Romay, G., Chirinos, D., Geraud-Pouey, F. & Desbiez, C. (2010). Association of an atypical alphasatellite with a bipartite New World begomovirus. *Arch Virol* 155, 1843-1847.
Rosario, K., Duffy, S. & Breitbart, M. (2012). A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch Virol* 157, 1851-1871.
Rosario, K., Padilla-Rodriguez, M., Kraberger, S., Stainton, D., Martin, D. P., Breitbart, M. & Varsani, A. (2013). Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Epiprocta) from Puerto Rico. *Virus Res* 171, 231-237.
Rybicki, E. P. (2014). A Top Ten list for economically important plant viruses. *Arch Virol*.
Rybicki, E. P. & Pietersen, G. (1999). Plant virus disease problems in the developing world. *Adv Virus Res* 53, 127-175.
Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. & Kojima, M. (1998). Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. *J Gen Virol* 79, 3111-3118.
Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G. & Stanley, J. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* 293, 63-74.
Savory, F. R. & Ramakrishnan, U. (2014a). Asymmetric patterns of reassortment and concerted evolution in Cardamom bushy dwarf virus. *Infection, Genetics and Evolution* 24, 15-24.
Savory, F. R. & Ramakrishnan, U. (2014b). Cryptic diversity and habitat partitioning in an economically important aphid species complex. *Infection, Genetics and Evolution*.
Savory, F. R., Varma, V. & Ramakrishnan, U. (2014). Identifying geographic hot spots of reassortment in a multipartite plant virus. *Evolutionary Applications* 7, 569-579.
Seguin, J., Rajeswaran, R., Malpica-López, N., Martin, R. R., Kasschau, K., Dolja, V. V., Otten, P., Farinelli, L. & Pooggin, M. M. (2014). De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLoS ONE* 9.
Selvarajan, R., Sheeba, M. M., Balasubramanian, V., Rajmohan, R., Dhevi, N. L. & Sasireka, T. (2010). Molecular Characterization of Geographically Different Banana bunchy top virus Isolates in India. *Indian Journal of Virology* 21, 110-116.
Shackelton, L. A., Parrish, C. R., Truyen, U. & Holmes, E. C. (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci U S A* 102, 379-384.
Sharman, M., Thomas, J. E., Skabo, S. & Holton, T. A. (2008). Abaca bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch Virol* 153, 135-147.
Shekhawat, U. K. S., Ganapathi, T. R. & Hadapad, A. B. (2012). Transgenic banana plants expressing small interfering RNAs targeted against viral replication initiation gene display high-level resistance to banana bunchy top virus infection. *J Gen Virol* 93, 1804-1813.
Shepherd, D. N., Dugdale, B., Martin, D. P., Varsani, A., Lakay, F. M., Bezuidenhout, M. E., Monjane, A. L., Thomson, J. A., Dale, J. & Rybicki, E. P. (2014). Inducible Resistance to Maize Streak Virus. *PLoS ONE* 9, e105932.
Shepherd, D. N., Martin, D. P., Lefeuvre, P., Monjane, A. L., Owor, B. E., Rybicki, E. P. & Varsani, A. (2008). A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *J Virol Methods* 149, 97-102.
Sicard, A., Yvon, M., Timchenko, T., Gronenborn, B., Michalakis, Y., Gutierrez, S. & Blanc, S. (2013). Gene copy number is differentially regulated in a multipartite virus. *Nat Commun* 4.
Simmonds, N. W. & Shepherd, K. (1955). The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society of London, Botany* 55, 302-312.
Simons, J. N. & Coe, D. M. (1958). Transmission of pseudo-curly top virus in Florida by a treehopper. *Virology* 6, 43-48.
Staginnus, C., Iskra-Caruana, M. L., Lockhart, B., Hohn, T. & Richert-Pöggeler, K. R. (2009). Suggestions for a nomenclature of endogenous pararetroviral sequences in plants. *Arch Virol* 154, 1189-1193.
Stanley, J. & Townsend, R. (1985). Characterisation of DNA forms associated with cassava latent virus infection. *Nucleic Acids Res* 13, 2189-2206.

Stavolone, L., Herzog, E., Leclerc, D. & Hohn, T. (2001). Tetramerization is a conserved feature of the virion-associated protein in plant pararetroviruses. *J Virol* 75, 7739-7743.

Stenger, D. C., Stevenson, M. C., Hormuzdi, S. G. & Bisaro, D. M. (1992). A number of subgenomic DNAs are produced following agroinoculation of plants with beet curly top virus. *J Gen Virol* 73, 237-242.

Stenzel, T., Piasecki, T., Chrząstek, K., Julian, L., Muhire, B. M., Golden, M., Martin, D. P. & Varsani, A. (2014). Pigeon circoviruses display patterns of recombination, genomic secondary structure and selection similar to those of beak and feather disease viruses. *J Gen Virol* 95, 1338-1351.

Streck, A., Bonatto, S. L., Homeier, T., Souza, C. K., Gonçalves, K. R., Gava, D., Canal, C. W. & Truyen, U. (2011). High rate of viral evolution in the capsid protein of porcine parvovirus. *J Gen Virol* 92, 2628-2636.

Su, H.-J., Wu, R.-Y. & Tsao, L.-Y. (1992). Ecology of banana bunchy-top virus disease. In *Proceedings of the International Symposium on Recent Developments in Banana Cultivation Technology, Los Baños, Philippines: INIBAP-ASPNET*, pp. 308-312.

Tentchev, D., Verdin, E., Marchal, C., Jacquet, M., Aguilar, J. M. & Moury, B. (2011). Evolution and structure of Tomato spotted wilt virus populations: Evidence of extensive reassortment and insights into emergence processes. *J Gen Virol* 92, 961-973.

Teycheney, P.-Y. & Geering, A. D. (2011). Endogenous viral sequences in plant genomes. *Recent Advances in Plant Virology*, 347-366.

Thomas, J. E. & Dietzgen, R. G. (1991). Purification, characterization and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *J Gen Virol* 72, 217-224.

Thomas, J. E., Geering, A. D. W., Gambley, C. F., Kessling, A. F. & White, M. (1997). Purification, properties, and diagnosis of banana bract mosaic potyvirus and its distinction from abaca mosaic potyvirus. *Phytopathology* 87, 698-705.

Thottappilly, G., Dahal, G. & Lockhart, B. E. L. (1998). Studies on a Nigerian isolate of banana streak badnavirus: I. Purification and enzyme-linked immunosorbent assay. *Ann Appl Biol* 132, 253-261.

Timchenko, T., De Kouchkovsky, F., Katul, L., David, C., Vetten, H. J. & Gronenborn, B. (1999). A single Rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *J Virol* 73, 10173-10182.

Timchenko, T., Katul, L., Aronson, M., Vega-Arreguín, J. C., Ramirez, B. C., Vetten, H. J. & Gronenborn, B. (2006). Infectivity of nanovirus DNAs: Induction of disease by cloned genome components of Faba bean necrotic yellows virus. *J Gen Virol* 87, 1735-1743.

Timchenko, T., Katul, L., Sano, Y., de Kouchkovsky, F., Vetten, H. J. & Gronenborn, B. (2000). The master Rep concept in nanovirus replication: Identification of missing genome components and potential for natural genetic reassortment. *Virology* 274, 189-195.

To, K. K. W., Chan, J. F. W., Chen, H., Li, L. & Yuen, K. Y. (2013). The emergence of influenza A H7N9 in human beings 16 years after influenza A H5N1: A tale of two cities. *The Lancet Infectious Diseases* 13, 809-821.

Tsegaye, A. & Struik, P. C. (2002). Analysis of enset (Ensete ventricosum) indigenous production methods and farm-based biodiversity in major enset-growing regions of southern Ethiopia. *Exp Agric* 38, 291-315.

Tushemereirwe, W., Kangire, A., Ssekiwoko, F., Offord, L. C., Crozier, J., Boa, E., Rutherford, M. & Smith, J. J. (2004). First report of Xanthomonas campestris pv. musacearum on banana in Uganda. *Plant Pathology* 53, 802.

van der Walt, E., Rybicki, E. P., Varsani, A., Polston, J. E., Billharz, R., Donaldson, L., Monjane, A. L. & Martin, D. P. (2009). Rapid host adaptation by extensive recombination. *J Gen Virol* 90, 734-746.

Varma, P. M. & Capoor, S. P. (1964). 'Foorkey' disease of large cardamom. *Indian J Agric Sci* 34, 56-62.

Varsani, A., Martin, D. P., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Murilo Zerbini, F. & Brown, J. K. (2014a). Revisiting the classification of curtoviruses based on genome-wide pairwise identity. *Arch Virol* 159, 1873-1882.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J., Murilo Zerbini, F. & Martin, D. (2014b). Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch Virol*, 1-11.

Vega-Rocha, S., Gronenborn, B., Gronenborn, A. M. & Campos-Olivas, R. (2007). Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circoviras structures. *Biochemistry* 46, 6201-6212.

Vishnoi, R., Kumar, S. & Raj, S. K. (2013). Molecular characterization of a Cucumber mosaic virus isolate associated with mosaic disease of banana in India. *Phytoparasitica* 41, 545-555.

Vishnoi, R., Raj, S. & Prasad, V. (2009). Molecular characterization of an Indian isolate of Banana bunchy top virus based on six genomic DNA components. *Virus Genes* 38, 334-344.

Wang, H. I., Chang, C. H., Lin, P. H., Fu, H. C., Tang, C. & Yeh, H. H. (2013). Application of Motif-Based Tools on Evolutionary Analysis of Multipartite Single-Stranded DNA Viruses. *PLoS ONE* 8.

Wanitchakorn, R., Hafner, G. J., Harding, R. M. & Dale, J. L. (2000a). Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *J Gen Virol* 81, 299-306.

Wanitchakorn, R., Harding, R. M. & Dale, J. L. (1997). Banana bunchy top virus DNA-3 encodes the viral coat protein. *Arch Virol* 142, 1673-1680.

Wanitchakorn, R., Harding, R. M. & Dale, J. L. (2000b). Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch Virol* 145, 593-602.

Watanabe, S. & Bressan, A. (2013). Tropism, compartmentalization and retention of banana bunchy top virus (Nanoviridae) in the aphid vector Pentalonia nigronervosa. *J Gen Virol* 94, 209-219.

Watanabe, S., Greenwell, A. & Bressan, A. (2013). Localization, Concentration, and Transmission Efficiency of Banana bunchy top virus in Four Asexual Lineages of Pentalonia aphids. *Viruses* 5, 758-776.

Wei, J., Liu, H., Liu, F., Zhu, M., Zhou, X. & Xing, D. (2014). Miniaturized paper-based gene sensor for rapid and sensitive identification of contagious plant virus. *ACS Applied Materials and Interfaces* 6, 22577-22584.

Wildy, P. (1971). *Classification and Nomenclature of Viruses*.

Wright, E. A., Heckel, T., Groenendijk, J., Davies, J. W. & Boulton, M. I. (1997). Splicing features in maize streak virus virion- and complementary-sense gene expression. *Plant J* 12, 1285-1297.

Wu, R.-Y., You, L.-R. & Soong, T.-S. (1994). Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus. *Phytopathology* 84, 952-957.

Wu, R. Y. & Su, H. J. (1990). Production of monoclonal antibodies against banana bunchy top virus and their use in enzyme-linked immunosorbent assay. *Journal of Phytopathology-Phytopathologische Zeitschrift* 128, 203-208.

Xie, O., Suarez-Lopez, P. & Gutierrez, C. (1995). Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: Requirement for efficient viral DNA replication. *EMBO Journal* 14, 4073-4082.

Xie, W. S. & Hu, J. S. (1995). Molecular cloning, sequence analysis, and detection of banana bunchy top virus in Hawaii. *Phytopathology* 85, 339-347.

Yazdi, H. R. B., Heydarnejad, J. & Massumi, H. (2008). Genome characterization and genetic diversity of beet curly top Iran virus: A geminivirus with a novel nonanucleotide. *Virus Genes* 36, 539-545.

Yu, N. T., Zhang, Y. L., Feng, T. C., Wang, J. H., Kulye, M., Yang, W. J., Lin, Z. S., Xiong, Z. & Liu, Z. X. (2012). Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* 44, 488-494.

Zaffalon, V., Mukherjee, S. K., Reddy, V. S., Thompson, J. R. & Tepfer, M. (2012). A survey of geminiviruses and associated satellite DNAs in the cotton-growing areas of northwestern India. *Arch Virol* 157, 483-495.

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W. & McKenna, R. (2001). Structure of the maize streak virus geminate particle. *Virology* 279, 471-477.

Zhou, X. (2013). Advances in understanding begomovirus satellites. In *Annu Rev Phytopathol*, pp. 357-381.

Zulperi, D. & Sijam, K. (2014). First report of Ralstonia solanacearum race 2 biovar 1 causing moko disease of banana in Malaysia. *Plant Dis* 98, 275.

Chapter 2

Evidence of inter-component recombination, intra-component recombination and reassortment in *Banana bunchy top virus*

Contents

This work has been published in Journal of General Virology and is presented in a similar manner to that of the publication:

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E.J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T.H., Aholelei,W., Taufa, L., Thomas, J.E, Collings, D.A., Martin, D.P., Varsani, A. (2012) Evidence of inter-component recombination, intra-component recombination and reassortment in Banana bunchy top virus. Journal of General Virology, 93, 1103-1119

2.1 Abstract

Banana bunchy top virus (BBTV; Family *Nanoviridae*, Genus *Babuvirus*) is a multi-component, single-stranded DNA virus (ssDNA), which causes widespread banana crop losses throughout tropical Africa and Australasia. We determined the full genome sequences of twelve BBTV isolates from the Kingdom of Tonga and analysed these together with previously determined BBTV sequences to show that reassortment and both inter- and intra-component recombination have all been relatively frequent occurrences during BBTV evolution. We find that whereas DNA-U3 components display evidence of complex inter- and intra-component recombination, all of the South Pacific DNA-R components have a common intra-component recombinant origin spanning the replication associated protein gene. Altogether, the DNA-U3 and DNA-M components display a greater degree of inter-component recombination than the DNA-R, -S, -C and -M components. The breakpoint distribution of the inter-component recombination events reveals a primary recombination hotspot around the 5' side of the common region major and, in accordance with recombination hotspots detectable in related single stranded DNA viruses, a secondary recombination hotspot near the origin of virion strand replication.

2.2 Introduction

Banana bunchy top disease (BBTD) only affects plants in the family *Musaceae* where it causes serious losses in banana (*Musa spp.*) in the Pacific, south and Southeast Asia, India and sub-Saharan Africa (Kagy *et al.*, 2001; Kumar *et al.*, 2011; Thomas *et al.*, 2000). Infection causes plant stunting, leaf bunching and severely impaired fruit development. Infected leaves also show dark green streak and hooking patterns on their undersides extending from the midrib out towards the leaf margins. Following the first description of BBTD in Fiji in 1889 (Magee *et al.*, 1927; Wardlaw, 1961), BBTD has now been reported from most banana growing regions in the old world, where it seriously impacts banana production (Dale, 1987).

Despite the early description of BBTD and the establishment of the disease's viral nature in 1927 (Magee *et al.*, 1927; Wardlaw, 1961), it was only in the early 1990s that BBTD's causal agent was identified as a novel ssDNA virus that was subsequently named BBTV (Burns *et al.*, 1995; Harding *et al.*, 1993). BBTV is the type member of the *Babuvirus* genus in the family *Nanoviridae*. *Abaca bunchy top virus* (ABTV) is currently the only other characterised *Babuvirus* species and it also infects *Musa spp*. BBTV is aphid-transmitted and has a multipartite genome which is generally considered to consist of six circular ssDNAs (each ~1.1 kilobase (kb) long) that are individually encapsidated within separate icosahedral virions (each ~18–20 nanometre in diameter). The *Babuvirus* genus is distinguished from the only other genus in the family *Nanoviridae*, the genus *Nanovirus*, by having only six components compared to up to eight components found in nanoviruses. Each of the six babuvirus genome components encodes a different protein in the virion sense strand. These include a replication-associated protein, (Rep; encoded on DNA-R), a protein with unknown function (encoded on DNA-U3), a capsid protein (CP; encoded on DNA-S), a movement protein (MP; encoded on DNA-M), a cell-cycle link protein (Clink; encoded on DNA-C) and a nuclear shuttle protein (NSP; encoded on DNA-N) (Beetham *et al.*, 1997; Burns *et al.*, 1995; Karan, 1997; Vetten *et al.*, 2005).

In addition to these various genes, there are two common regions conserved across all six components (Burns *et al.*, 1995). These are the common region major (CR-M) and the common region stem-loop (CR-SL). Although the function of CR-M remains less clear, it contains sequence domains that are commonly associated with promoters and may therefore be involved in transcriptional regulation (Burns *et al.*, 1995). The CR-M also contains a

primer binding site for the initiation of complementary strand synthesis (Hafner *et al.*, 1997). The CR-SL is a sequence capable of forming a hairpin structure that contains the highly conserved nonanucleotide, TATTATTAC, within its loop and repeated sequences that are probably Rep sequence-specific binding motifs (Burns *et al.*, 1995). Based on a similarly structured region in geminivirus genomes, it is hypothesized that the CR-SL contains the origin of BBTV virion strand replication (Hanley-Bowdoin *et al.*, 1999).

BBTV and ABTV, the two distinct babuviruses, share ~54-76% pairwise nucleotide identity across all their components, with their CP and Rep proteins sharing ~80% and 85-87% amino acid sequence identity, respectively (Sharman *et al.*, 2008). Despite numerous individual BBTV component sequences having been deposited in public databases, few complete genomes (containing all six components) are available (Figure 2.1a).

Based on the phylogenetic relationships amongst DNA-R component sequences, Karan *et al.*, (1994) designated two different BBTV lineages: the South Pacific Group (SPG), which also included Indian, and African isolates, and the Asian group (AG). The timescales over which these lineages diverged are unknown and might span the history of banana cultivation (Karan *et al.*, 1994; Perrier *et al.*, 2011). Recently, however, it was demonstrated that the nanovirus, *Faba bean necrotic stunt virus* (FBNSV), has the capacity to evolve at a rate of 1.78×10^{-3} substitution/site/year (Grigoras *et al.*, 2010). This is similar to geminiviruses, another family of plant ssDNA viruses (Harkins *et al.*, 2009; van der Walt *et al.*, 2008), but is far higher than that of double-stranded DNA viruses (Duffy *et al.*, 2008). This very high basal mutation rate, and evidence of frequent homologous recombination both between different components of the same genome (Hu *et al.*, 2007; Hughes, 2004; Lefeuvre *et al.*, 2009) and between homologous components in different genomes (Hu *et al.*, 2007; Hughes, 2004; Hyder *et al.*, 2011), suggests that the two main BBTV lineages could have diverged recently - i.e. only hundreds of years ago rather than millennia.

It is reasonable to suspect that the ancestral BBTV variant that first caused BBTD originated from Southeast Asia and the Pacific region where bananas originated (Perrier *et al.*, 2011). This region includes the Southeast Asian peninsula, the Philippines, Indonesia, New Guinea, the Pacific Islands and Australia. Since BBTV may have evolved from this area, the South Pacific region is expected to harbour greater BBTV diversity than has been seen in the remainder of the world. Despite the early documentation of BBTD in the South Pacific region, the only complete BBTV genome (comprising all six components) sequence that has

ever been determined from this area is an isolate from Australia. In addition, only four other full BBTV genomes have been sequenced worldwide (one from India, two from China, one from Taiwan). Thus, previous phylogenetic analyses that have been performed on BBTV have used individual BBTV component sequences rather than full genomes.

Although sequence analyses of individual BBTV components have revealed geographical clustering of BBTV isolates and some evidence of intra- and inter-component recombination (Hu *et al.*, 2007; Hyder *et al.*, 2011; Karan *et al.*, 1994; Karan, 1997; Sharman *et al.*, 2008; Timchenko *et al.*, 2006b; Wanitchakorn *et al.*, 2000; Xie & Hu, 1995), we realised that analyses of full genome sequences from more South Pacific region BBTV isolates should provide a clearer view both of these evolutionary processes and of genome component reassortment. Here we report the complete genome sequences of twelve BBTV isolates from the Kingdom of Tonga in the Southwest Pacific Ocean, approximately 1600 km northeast of New Zealand. In order to assess the relative roles of homologous recombination and reassortment in BBTV evolution we have analysed these together with all complete BBTV genomes and individual component sequences found in public databases.

2.3 Materials and methods

2.3.1 Sample collection and virus isolates

In 2010, leaves from BBTD symptomatic plants were collected from the three major islands of the Kingdom of Tonga (Tongatapu n=6, Ha'apai n=3, Vava'u n=3). Plant material was collected by the authors DS, SK, MW, EJW, MH, SL, IK, THF, WA, LT, DAC, AV and students from Tonga College Tongatapu. Total DNA was extracted from these using the DNeasy Plant Mini Kit (Qiagen, USA).

Viral DNA was amplified by rolling circle amplification (RCA) using the Illustra TempliPhi Amplification Kit (GE Healthcare, USA) as described previously (Owor *et al.*, 2007; Shepherd *et al.*, 2008). The amplified concatemers were digested using *Xba*I, *Sma*1 and *Hind*III restriction enzymes. The resulting ~1.1 kb fragments were ligated to appropriately digested pUC19 and sequenced at Macrogen (Korea). Based on the resulting sequences, coupled with additional BBTV sequences available in GenBank, we designed back to back primers for specific amplification of each full component. DNA-R: Forward 5'-TTG AGA AAC GAA AGG GRA GC-3', Reverse 5'-GGT GTG CGC CTG GGA AG-3'. DNA-U3: Forward 5'-WWT TAA TTC GTA GGA CAC GTG GAC G-3, Reverse 5'-WWT TAA TTC

GTG TWK CTT GCT GCG C-3'. DNA-S: Forward 5'-GGT TCC TGA AAA YAC CGT C-3', Reverse 5'-AAT ATT GAY CCT ARM GAC GAG YAG TC-3'. DNA-M: Forward 5'-GAA TKG TTT YTG TTY WTY GVA GC-3', Reverse 5'-RAA GAA TAG TTT MAC CCG CTC-3'. DNA-C: Forward 5'-GCG AAT AYY TGA AGA AAC CAT G-3', Reverse 5'-TAT ACG AAT RAT AGA AGT CTT CAW AYC-3'. DNA-N: Forward 5'-TCA AGA CMT GTA CYC ATG G-3', Reverse 5'-RTT GTG ATT CYG CCC AAT CC-3'

DNA-R, -M -C, and -N were all amplified using the same protocol 94°C [2 min] (94°C [30 s], 50°C [30 s], 72°C [1:15 min]) x 25 cycles, 72°C [1 min], 4°C [10 min]. DNA-U3 and DNA-S were amplified with a similar protocol but with annealing temperatures of 53°C and 48°C respectively. The PCR products were cloned and sequenced (Macrogen). In some samples, where components were difficult to amplify by PCR alone, RCA was use to enrich viral DNA prior to PCR. Sequences were assembled and manually edited where necessary using DNAMAN (version 5.2.9; Lynnon Biosoft) and MEGA5 (Tamura *et al.*, 2011). For each sample we obtained a single sequence of each of the six components.

2.3.2 Phylogenetic analyses

All full length sequences of BBTV available in GenBank (Table 2.1) were downloaded on the 12th of March 2011 and compared with all six components of the twelve Tongan isolates. Two full-length ABTV sequences for each component were also downloaded as outgroups. Sequences were grouped and analysed based on the components, each beginning at the origin of replication (TATTACC). All maximum-likelihood (ML) phylogenetic trees were constructed using PhyML (Guindon & Gascuel, 2003) with the best fit model selected by RDP4 (Martin *et al.*, 2010). Pairwise distance (p-distance) comparisons (determined with pairwise deletion of alignment gaps) were carried out using MEGA 5 (Tamura *et al.*, 2011).

2.3.3 Analysis of recombination and reassortment

Evidence of recombination was sought using the RDP (Martin & Rybicki, 2000), GENECONV (Padidam *et al.*, 1999), BOOTSCAN (Martin *et al.*, 2005), MAXCHI (Smith, 1992), CHIMAERA (Posada & Crandall, 2001), SISCAN (Gibbs *et al.*, 2000), and 3SEQ (Boni *et al.*, 2007) methods implemented in the program RDP4 (version 4.13; (Martin *et al.*, 2010)). Only potential recombination signals detected by at least three recombination detection methods, coupled with phylogenetic evidence of recombination, were considered significant evidence of the signals representing genuine recombination events.

The separate components of BBTV and ABTV genomes were aligned individually (DNA-R, n=76; DNA-U3, n=33; DNA-S, n=48; DNA-M, n=27; DNA-C, n=27; DNA-N, n=26) and the datasets were labelled by component (R, U3, S, M, C, N). The individual components (IC datasets) were independently analysed for recombination (Martin *et al.*, 2010). In addition to these six datasets, we generated two datasets, both containing all 237 available BBTV genome component sequences, including the two ABTV variants. The first dataset, called the single component (SC) dataset contained all aligned component sequences linearised at the CR-SL region in a single alignment. The second dataset, called the concatenated (CON) dataset, consisted of all 237 sequences constructed by concatenating cognate genome component sequences in the order DNA- R, -U3, -S, -M, -C, -N. Each component started with the virion strand origin hairpin-loop sequence TATTACC. While some of the concatenated sequences contained all six components others contained only a single component with alignment gaps corresponding to unsequenced components. Recombination analysis of both the SC and CON datasets was performed using RDP4 (Martin *et al.*, 2010) as outlined above. Evidence within the SC dataset of inter-component recombination breakpoints clustering into recombination hotspots was statistically tested using the approach described by Heath *et al.* (2006).

Whereas recombination between homologous components was identifiable using both datasets, our intention was to detect reassortment events with the CON dataset by identifying these as recombination events with estimated recombination breakpoint-pairs falling at the interfaces between concatenated components. With the SC dataset we sought to identify inter-component recombination within genomes. To check whether recombination events detected with the concatenated dataset were also represented in the single component dataset, the recombination events detected using this dataset were mapped onto a reordered alignment with all components represented as separate sequences and all starting at the TATTACC sequences within the hairpins of their viron strand origins of replication.

Table 2.1: Details of concatamers and the GenBank accession numbers of all the sequences used in these analyses. Phylogenetic group is given SPG-South Pacific Group, AG-Asian Group.

ID	Isolate	Country	Group	DNA-R	DNA-U3	DNA-S	DNA-M	DNA-C	DNA-N
1-PK	Thatha	Pakistan	SPG	AM418538	-	AM418566	AM418541	AM418569	AM418568
2-PK	Tandojam	Pakistan	SPG	AM418536	-	AM418540	-	AM418564	-
3-IN	Lucknow	India	SPG	DQ256267	EU402601	EF687856	EU516323	EU051379	EU391633
4-AU	Australia Nambour	Australia	SPG	S56276	L41576	L41574	L41575	L41578	L41577
5-TW	TW3	Taiwan	AG	EU366169	EU366170	EU366171	EU366172	EU366173	-
6-US	Hawaiian	USA	SPG	U18077	U18078	-	-	U18079	-
7-JP	JN4	Japan	AG	AB108452	-	AB108449	-	-	-
8-JP	JK3	Japan	AG	AB108453	-	AB108450	-	-	-
9-JP	JY1	Japan	AG	AB108456	-	AB108451	-	-	-
10-PH	bp5	Philippines	AG	AB189067	-	AB189068	-	-	-
11-VN	V6	Vietnam	AG	AB113659	-	AB113661	-	-	-
12-VN	V14	Vietnam	AG	AB113660	-	AB113662	-	-	-
13-PH	aP32	Philippines	AG	AB250953	-	AB250956	-	-	-
14-PH	aP34	Philippines	AG	AB250954	-	AB250957	-	-	-
15-PH	bP26	Philippines	AG	AB250955	-	AB250958	-	-	-
16-ID	IG33	Indonesia	AG	AB186924	-	AB186927	-	-	-
17-ID	IG64	Indonesia	AG	AB186925	-	AB186928	-	-	-
18-ID	IJs11	Indonesia	AG	AB186926	-	AB186929	-	-	-
19-CN	NSP	China	AG	AF238875	-	AF238877	-	-	-
20-CN	NS	China	AG	AF238874	-	AF238876	-	-	-
21-CN	Hainan	China	AG	AY450396	AY606084	AY494786	AY494788	AY606085	AY494787
22-IN	Robusta BT-1	India	SPG	-	AY960129	-	AY948439	-	AY948438
23-CN	Haikou	China	AG	FJ463042	FJ463043	FJ463044	FJ463045	FJ463046	FJ463047
24-TW	V-1	Taiwan	AG / SPG	EF095161	EF095163	EF095164	EF095165	EF095166	-
25-TW	Severe phenotype	Taiwan	AG	DQ826390	DQ826391	DQ826393	DQ826394	DQ826395	DQ826396
26-PK	TJ1	Pakistan	SPG	AY996562	-	-	-	-	EF529519
27-IN	Kerala 1	India	SPG	FJ009238	FJ009239	-	-	-	-
28-IN	Kerala 2	India	SPG	FJ009240	EU140341	-	-	-	-
29-MM	MY01	Myanmar	SPG	AB252639	-	AB252642	-	-	-
30-MM	MY02	Myanmar	SPG	AB252640	-	AB252643	-	-	-
31-MM	MY03	Myanmar	SPG	AB252641	-	AB252644	-	-	-
32-FJ	Fiji	Fiji	SPG	AF416466	-	AF148944	-	-	-
33-IN	Tamil Nadu	India	SPG	EU140342	-	-	EU190971	EU190969	EU190970
34-EG	Egypt Kalubia	Egypt	SPG	AF102780	-	-	AF102783	-	-
35-TO	TO166 – Tonga'tapu	Tonga	SPG	JF957628	JF957640	JF957652	JF957664	JF957676	JF957688
36-TO	TO114 – Tonga'tapu	Tonga	SPG	JF957625	JF957637	JF957649	JF957661	JF957673	JF957685
37-TO	TO121 – Tonga'tapu	Tonga	SPG	JF957626	JF957638	JF957650	JF957662	JF957674	JF957686
38-TO	TOS28 – Tonga'tapu	Tonga	SPG	JF957636	JF957648	JF957660	JF957672	JF957684	JF957696
39-TO	TOS12 – Tonga'tapu	Tonga	SPG	JF957635	JF957647	JF957659	JF957671	JF957683	JF957695
40-TO	TO124 – Tonga'tapu	Tonga	SPG	JF957627	JF957639	JF957651	JF957663	JF957675	JF957687
41-TO	TO306 – Ha'apai	Tonga	SPG	JF957632	JF957644	JF957656	JF957668	JF957680	JF957692
42-TO	TO314 – Ha'apai	Tonga	SPG	JF957634	JF957646	JF957658	JF957670	JF957682	JF957694
43-TO	TO310 – Ha'apai	Tonga	SPG	JF957633	JF957645	JF957657	JF957669	JF957681	JF957693
44-TO	TO208 – Vava'u	Tonga	SPG	JF957629	JF957641	JF957653	JF957665	JF957677	JF957689
45-TO	TO224 – Vava'u	Tonga	SPG	JF957630	JF957642	JF957654	JF957666	JF957678	JF957690
46-TO	TO290 – Vava'u	Tonga	SPG	JF957631	JF957643	JF957655	JF957667	JF957679	JF957691
Abaca1	Q767	Malaysia	-	EF546813	EF546809	EF546810	EF546811	EF546812	EF546808
Abaca2	Q1108	Philippines	-	EF546807	EF546803	EF546804	EF546805	EF546806	EF546802
AM418537	Chambar	Pakistan	SPG	AM418537	-	-	-	-	-
AM418539	Hala	Pakistan	SPG	AM418539	-	-	-	-	-
AM418534	Nawabshah	Pakistan	SPG	AM418534	-	-	-	-	-
AM418535	Sakrand	Pakistan	SPG	AM418535	-	-	-	-	-

AB108455	JM6	Japan	AG	AB108455	-	-	-	-	-
AB108454	JM5	Japan	AG	AB108454	-	-	-	-	-
AB108457	JY3	Japan	AG	AB108457	-	-	-	-	-
AB108458	JY7	Japan	AG	AB108458	-	-	-	-	-
AF416472	Son La region	Vietnam	AG	AF416472	-	-	-	-	-
AF416473	Dien Bien Phu region	Vietnam	AG	AF416473	-	-	-	-	-
AF416474	Bac Ninh region	Vietnam	AG	AF416474	-	-	-	-	-
AF416475	Hue region	Vietnam	AG	AF416475	-	-	-	-	-
AF416479	Yen Bai region	Vietnam	AG	AF416479	-	-	-	-	-
AF416476	Buon Ma Thout region	Vietnam	AG	AF416476	-	-	-	-	-
AF416477	Da Nang region	Vietnam	AG	AF416477	-	-	-	-	-
AF416478	Ho Chi Minh City region	Vietnam	AG	AF416478	-	-	-	-	-
EF095162	V-1 clone a	Taiwan	SPG	EF095162	-	-	-	-	-
AF416467	Tonga 1994	Tonga	SPG	AF416467	-	-	-	-	-
AF416465	Egypt 1994	Egypt	SPG	AF416465	-	-	-	-	-
AF416469	Philippines 1994	Philippines	AG	AF416469	-	-	-	-	-
AF416470	India 1994	India	SPG	AF416470	-	-	-	-	-
AF416464	Vietnam 1994	Vietnam	AG	AF416464	-	-	-	-	-
AF110266	Zhangzhou 1998	China	AG	AF110266	-	-	-	-	-
AF246123	Guangdong-1	China	AG	AF246123	-	-	-	-	-
GQ374514	Zhangjiang	China	AG	GQ374514	-	-	-	-	-
DQ656118	Kanpur	India	SPG	DQ656118	-	-	-	-	-
DQ656119	Etawah	India	SPG	DQ656119	-	-	-	-	-
U97525	Chinese isolate C4	China	AG	U97525	-	-	-	-	-
AF416468	Taiwan 1994	Taiwan	AG	AF416468	-	-	-	-	-
AY996563	KHI	Pakistan	SPG	-	AY996563	-	-	-	-
EU046323	Bangalore	India	SPG	-	EU046323	-	-	-	-
GQ214699	TJ1	Pakistan	SPG	-	GQ214699	-	-	-	-
FJ859748	GH1	Pakistan	SPG	-	FJ859748	-	-	-	-
FJ859749	JS1	Pakistan	SPG	-	FJ859749	-	-	-	-
FJ773283	TW3	Taiwan	AG	-	FJ773283	-	-	-	-
DQ826392	Severe strain clone b	Taiwan	AG	-	DQ826392	-	-	-	-
AM418565	Kisanamari	Pakistan	SPG	-	-	AM418565	-	-	-
AM418567	Nasarpur	Pakistan	SPG	-	-	AM418567	-	-	-
GQ249344	Cameroon	Cameroon	SPG	-	-	GQ249344	-	-	-
AF148943	Burundi	Burundi	SPG	-	-	AF148943	-	-	-
AF148945	Vietnam 1999	Vietnam	AG	-	-	AF148945	-	-	-
AF148068	Philippines 1999	Philippines	AG	-	-	AF148068	-	-	-
AF148942	Taiwan 1999	Taiwan	AG	-	-	AF148942	-	-	-
AY953429	TN	India	SPG	-	-	-	AY953429	-	-
AF349568	Zhangzhou 2001	China	AG	-	-	-	AF349568	-	-
AY266417	NS	China	AG	-	-	-	-	AY266417	-
AY264347	NSP	China	AG	-	-	-	-	AY264347	-
AF238878	NS	China	AG	-	-	-	-	-	AF238878
AF238879	NSP	China	AG	-	-	-	-	-	AF238879
EF470243	NSP	China	AG	-	-	-	-	-	EF470243

2.4 Results and discussion

Only a few disparate genome components had been sequenced from South Pacific BBTV isolates. These include individual genome components from Tonga (DNA-R; (Karan *et al.*, 1994), Fiji (DNA-R and DNA-S; (Karan *et al.*, 1994), (Wanitchakorn *et al.*, 2000) and Hawaii (DNA-R, DNA-C and DNA-U3; (Xie & Hu, 1995). We determined the sequences of all six components from twelve Tongan BBTV isolates and compared them with all known BBTV sequences available in GenBank (Figure 2.1A, Table 2.1). Alphasatellites are frequently associated with nanoviruses (Horser *et al.*, 2001a; Horser *et al.*, 2001b; Rohde *et al.*, 1990). For this study we did not specifically look for any alphasatellites nor did we come across any in our sequence analysis.

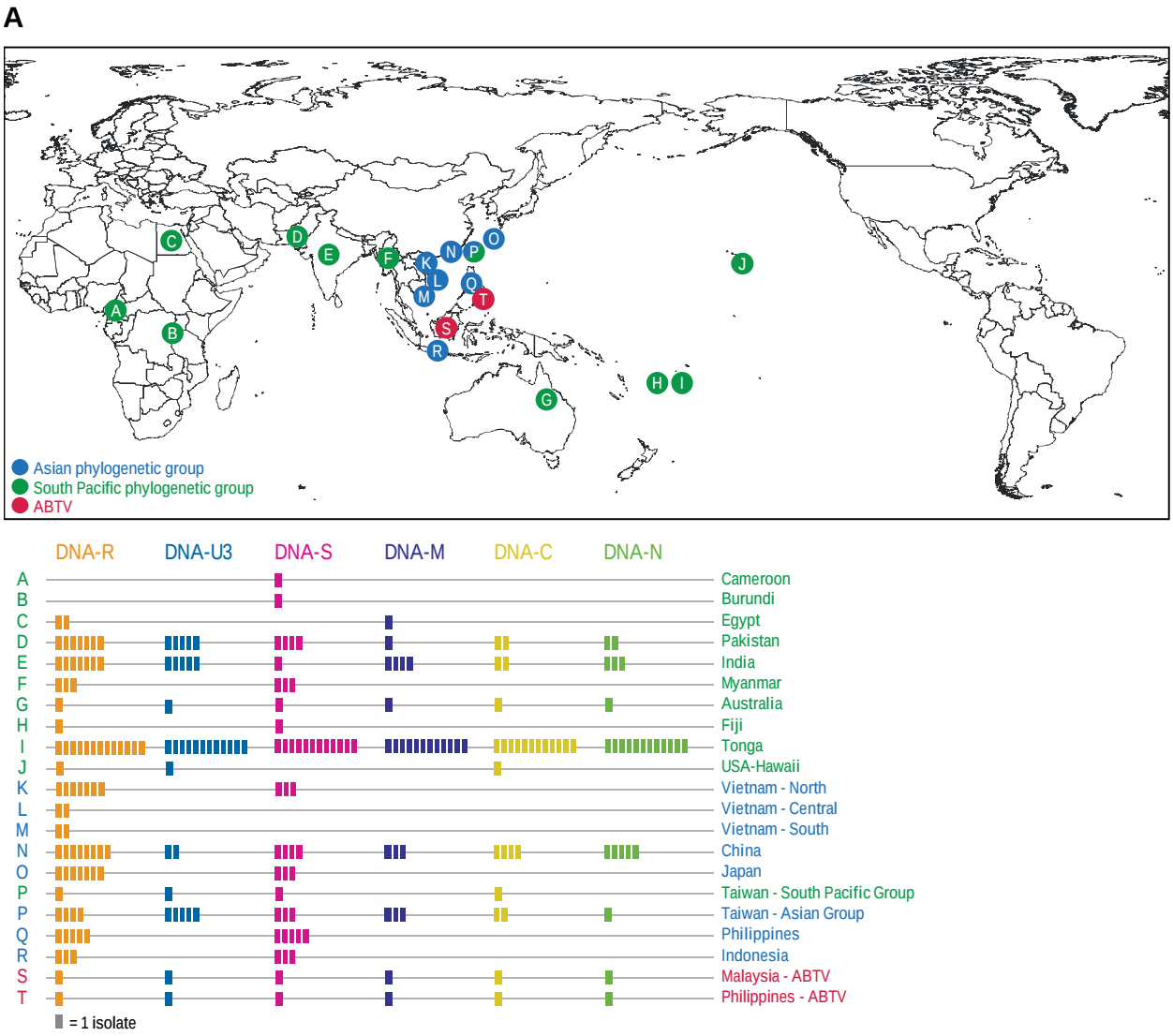


Figure 2.1: See next page for figure legend

Figure 2.1: A. Geographical locations of BBTV isolate component sequences that are currently available in GenBank and those determined in this study from the Kingdom of Tonga. Green circles depict isolates which fall within the South Pacific group, blue circles depict those within the Asian group. Red circles depict ABTV isolates. The numbers of component sequences that have been determined are depicted using bars (i.e. one bar represents one sequence) and are shown for each country. B. Inter-component recombination break point distribution plot indicating inter-component recombination hotspots (where the black line emerges above the grey areas). Whereas the dark grey area indicates the 95% probability interval of breakpoint clustering under the expectation that recombination breakpoints are randomly distributed, the light grey area indicates the corresponding 99% probability interval (Heath *et al.*, 2006). The locations of individual breakpoints are indicated by vertical lines above the plot.

2.4.1 Phylogenetic analysis

Two BBTV phylogenetic clades the SPG and the AG, have previously been identified based on analysis of DNA-R component sequences from ten different countries (Karan *et al.*, 1994). Our ML phylogenetic analyses of the six components confirmed both the separation of all available sequences into the same SPG and AG clades, and that, as expected, all Tongan sequences fall within the SPG (Figures 2.2 to 2.6).

Component by component, the Tongan DNA-R, DNA-U3, DNA-S, DNA-C, and DNA-N sequences share >96% identity with one another, while their DNA-M sequences are more diverse, sharing >93% identity with one another. A global analysis of all available BBTV component sequences indicates that (in order of degrees of conservation) BBTV DNA-R, DNA-S, DNA-C, DNA-N, DNA-M and DNA-U3 sequences share more than 88%, 87%, 85%, 84%, 82% and 73% sequence identity, respectively. Since the majority of the BBTV sequences available in GenBank are those of DNA-R (n= 74 including the twelve new Tongan sequences), the notion that DNA-R is the most conserved component is well supported.

AG BBTV isolates have predominantly been found in Southeast Asia and the Philippines, whereas the SPG has a broader tropical distribution spanning half the globe from Cameroon, Egypt, and Burundi in the west to Hawaii in the East (Figure 2.1A). The global distribution of BBTV has been artificially expanded by the trade and transport of infected propagules and aphids to regions outside its vector's normal range. Thus the distribution of the AG BBTV isolates occurs across the natural geographical range of *Musa acuminata* and *Musa balbisiana* whereas the SPG isolates occur across only the *M. balbisiana* range (Figure 1.1 shows the natural geographical range of wild *M. acuminata* and *M. balbisiana*). Unfortunately, basic replication and infectivity assays to definitively determine whether viruses in the two groups are differentially adapted to the two different banana species have not been carried out. Therefore Koch's postulates have not been fully fulfilled and hence the genomes presented here and in previous BBTV studies cannot be classified as complete infectious genomes. Due to the difficulty in initiating infections from cloned genomes, replication and infectivity has only recently been achieved for two virus species in the genus *Nanovirus* (Grigoras *et al.*, 2009; Timchenko *et al.*, 2006a).

Interestingly, the greatest diversity observed between BBTV isolates has been reported from Vietnam and Taiwan, which lie within the approximate area of the proposed original distribution of *M. balbisiana* (Jones, 1999; Perrier *et al.*, 2011). Indeed, an isolate containing a possible reassortment of AG and SPG components has been previously reported from Taiwan (Hu *et al.*, 2007).

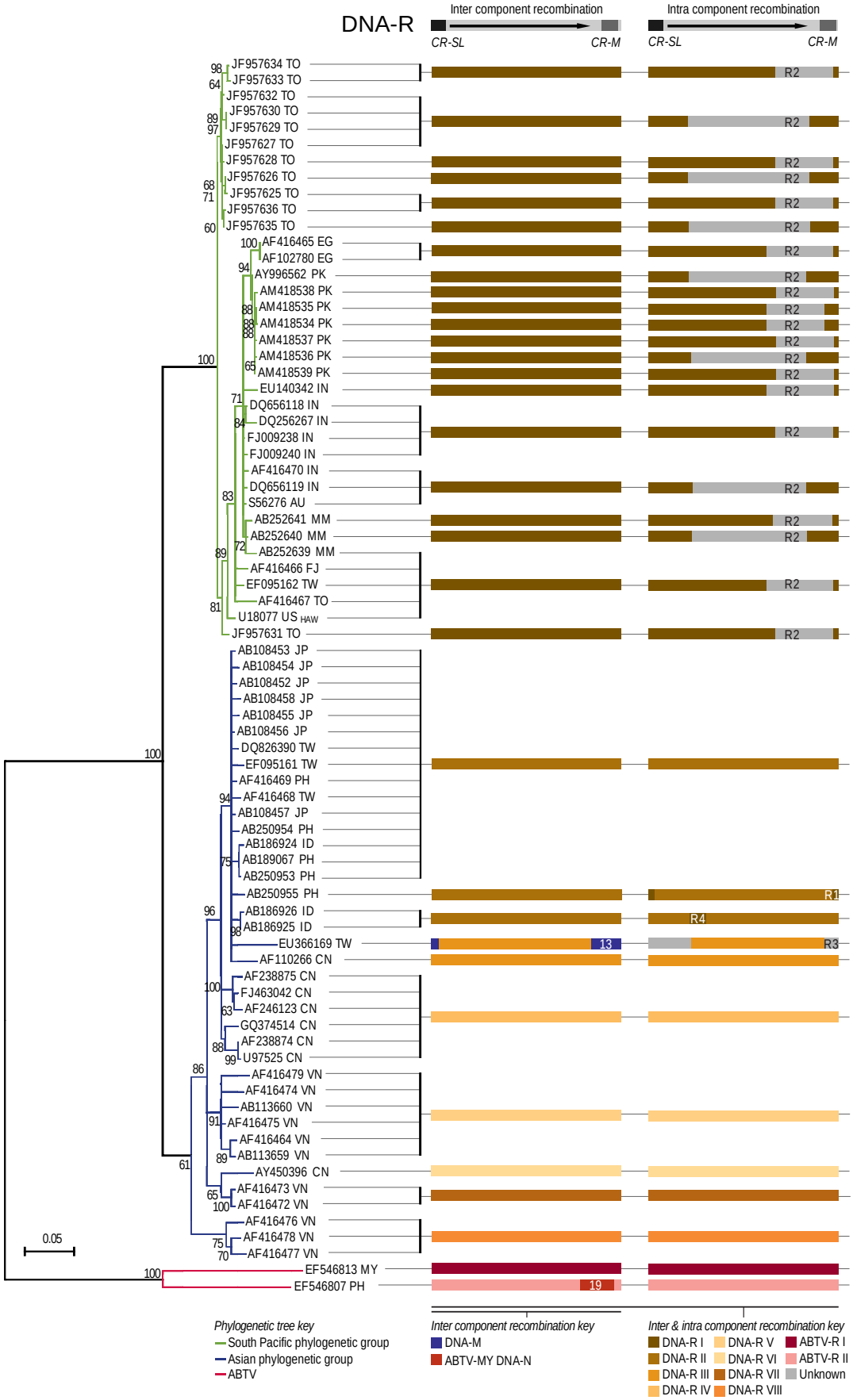


Figure 2.2: See next page for figure legend

Figure 2.2: Midpoint rooted maximum likelihood phylogenetic tree (nucleotide substitution model TN93+G4) of BBTV DNA-R components. Bootstrap support (1000 bootstrap replicates) is shown for each branch with >60% support. The South Pacific and Asian phylogenetic group sequences are indicated by green and blue branches, respectively. Two letter international country codes are provided with GenBank accession numbers. Sequences identified as either inter- or intra-component recombinants by RDP4 are shown with their corresponding breakpoint positions. Details of recombination break points and statistical support for evidence of recombination are provided in Supplementary Tables 2.1 and 2.2. Note that the intra-component recombination event labelled R3 (with an "unknown" parental sequence) is likely conflated with the inter-component recombination event labelled 13.

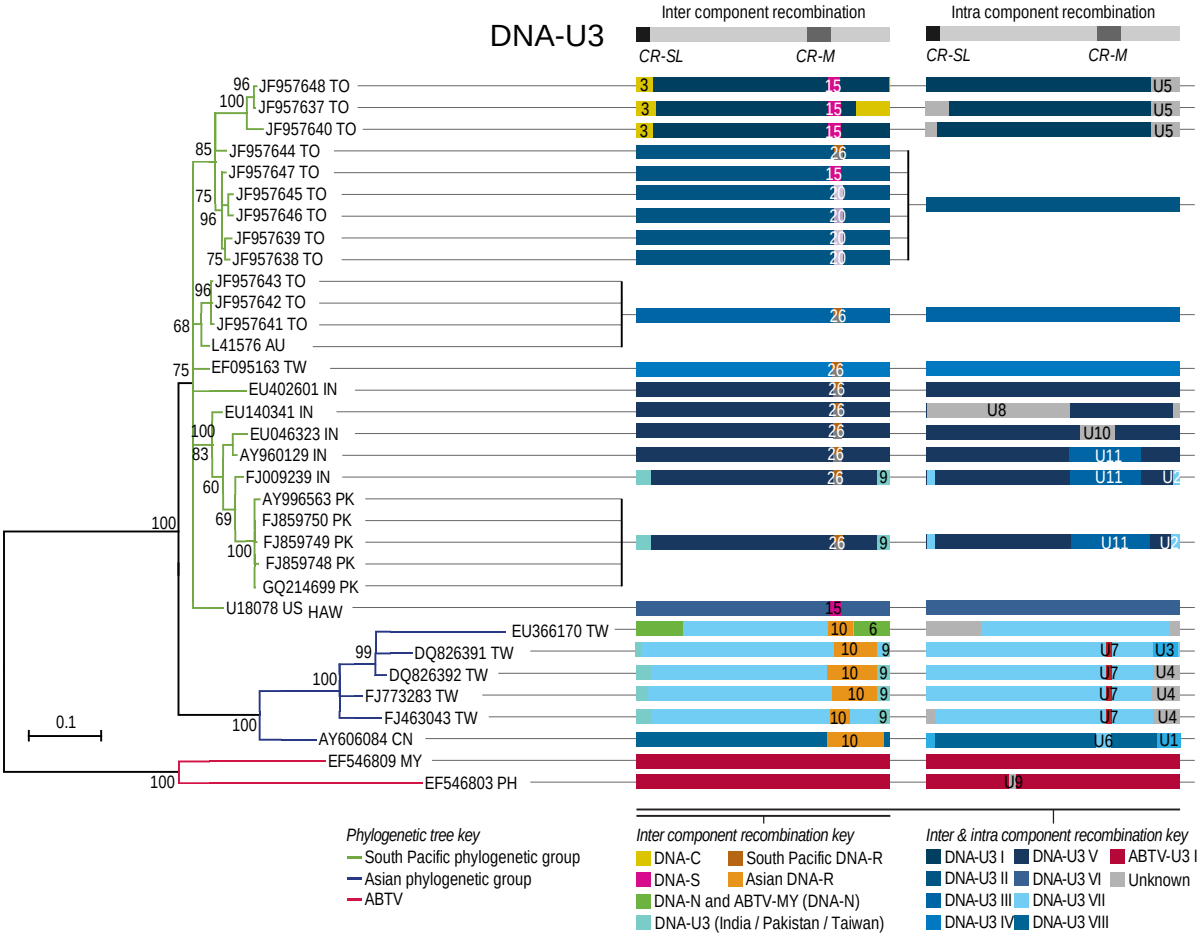


Figure 2.3: Midpoint rooted maximum likelihood phylogenetic tree (nucleotide substitution model GTR+G4) of BBTV DNA-U3 components. Bootstrap support (1000 bootstrap replicates) is shown for each branch with >60% support. The South Pacific and Asian phylogenetic group sequences are indicated by green and blue branches, respectively. ABTV isolates are indicated in red. Two letter international country codes are provided with GenBank accession numbers. Sequences identified as either inter- or intra-component recombinants by RDP4 are shown with their corresponding breakpoint positions. Details of recombination break points and statistical support for evidence of recombination are provided in Supplementary Tables 2.1 and 2.2. Note that the intra-component recombination event labelled U5 (with an "unknown" parental sequence) is likely conflated with the inter-component recombination event labelled 3. Also, the intra-component recombination events labelled U2, U3 and U4 are all likely conflated with the inter-component recombination event labelled 9.

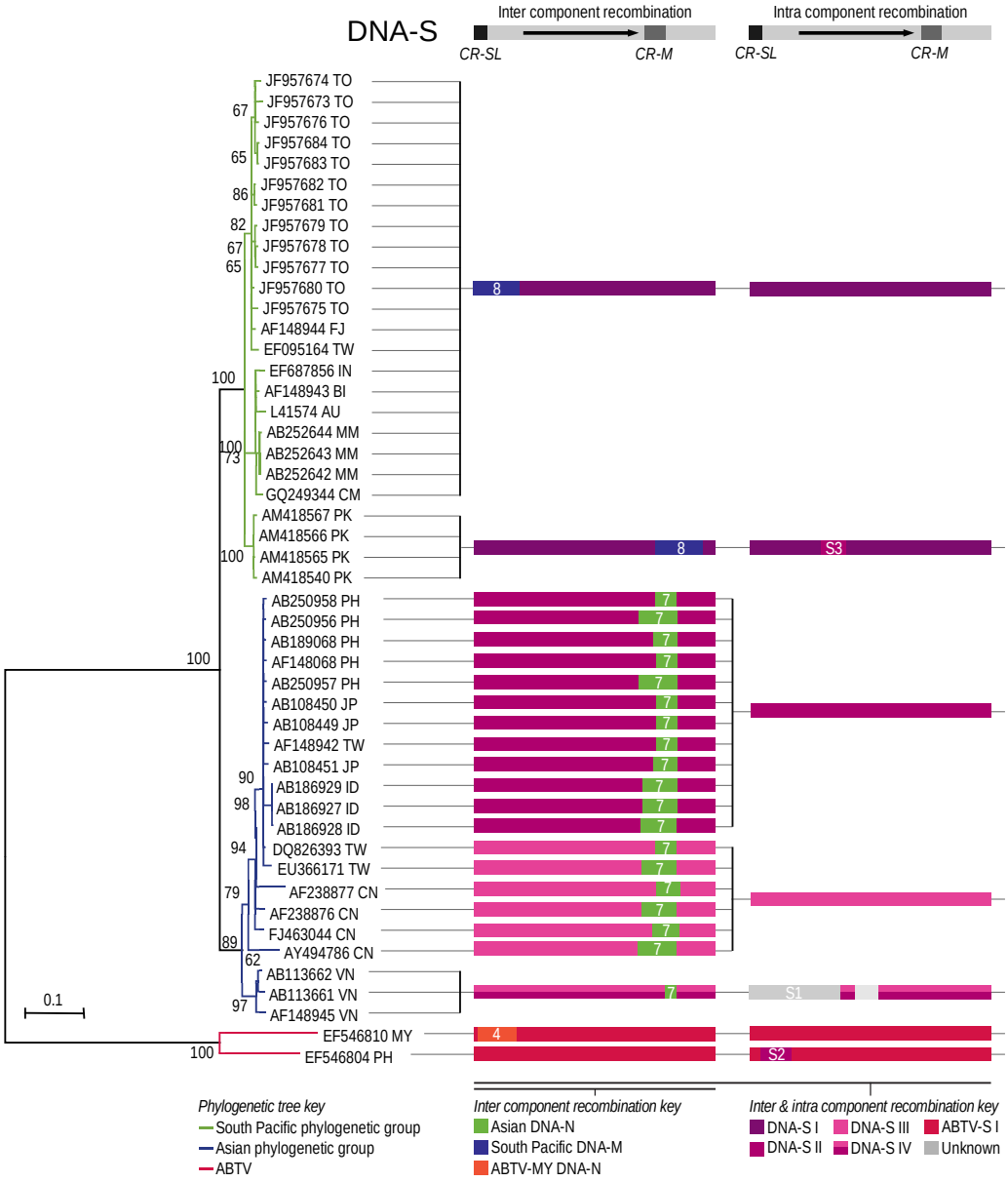


Figure 2.4: Midpoint rooted maximum likelihood phylogenetic tree (nucleotide substitution model TN93+G4) of BBTV DNA-S components. Bootstrap support (1000 bootstrap replicates) is shown for each branch with >60% support. The South Pacific and Asian phylogenetic group sequences are indicated by green and blue branches, respectively. ABTV isolates are indicated in red. Two letter international country codes are provided with GenBank accession numbers. Sequences identified as either inter- or intra-component recombinants by RDP4 are shown with their corresponding breakpoint positions. Details of recombination break points and statistical support for evidence of recombination are provided in Supplementary Tables 2.1 and 2.2.

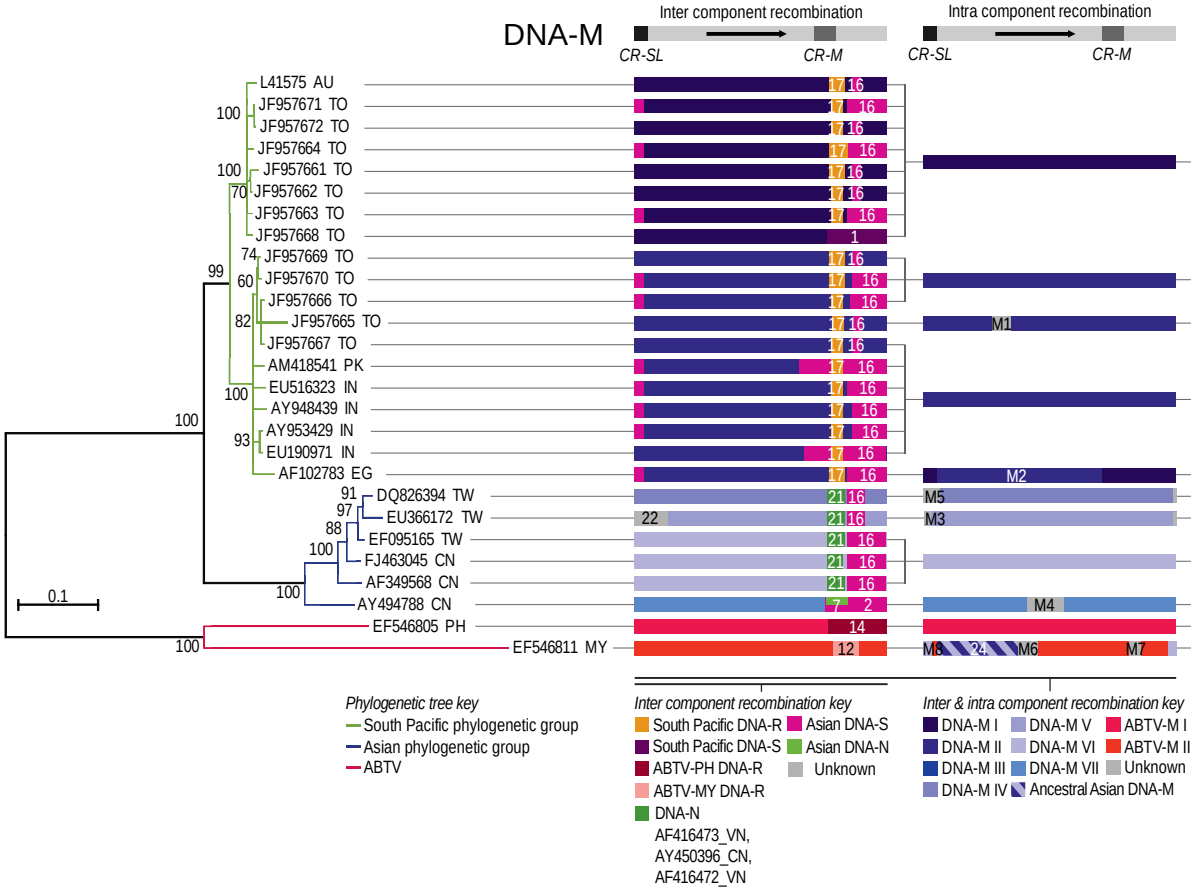


Figure 2.5: Midpoint rooted maximum likelihood phylogenetic tree (nucleotide substitution model TN93+G4) of BBTV DNA-M components. Bootstrap support (1000 bootstrap replicates) is shown for each branch with >60% support. The South Pacific and Asian phylogenetic group sequences are indicated by green and blue branches, respectively. ABTV isolates are indicated in red. Two letter international country codes are provided with GenBank accession numbers. Sequences identified as either inter- or intra-component recombinants by RDP4 are shown with their corresponding breakpoint positions. Details of recombination break points and statistical support for evidence of recombination are provided in Supplementary Tables 2.1 and 2.2. Note that the intra-component recombination events labelled M3 and M7 (with "unknown" parental sequences) are likely respectively conflated with the inter-component recombination events labelled 22 and 12.

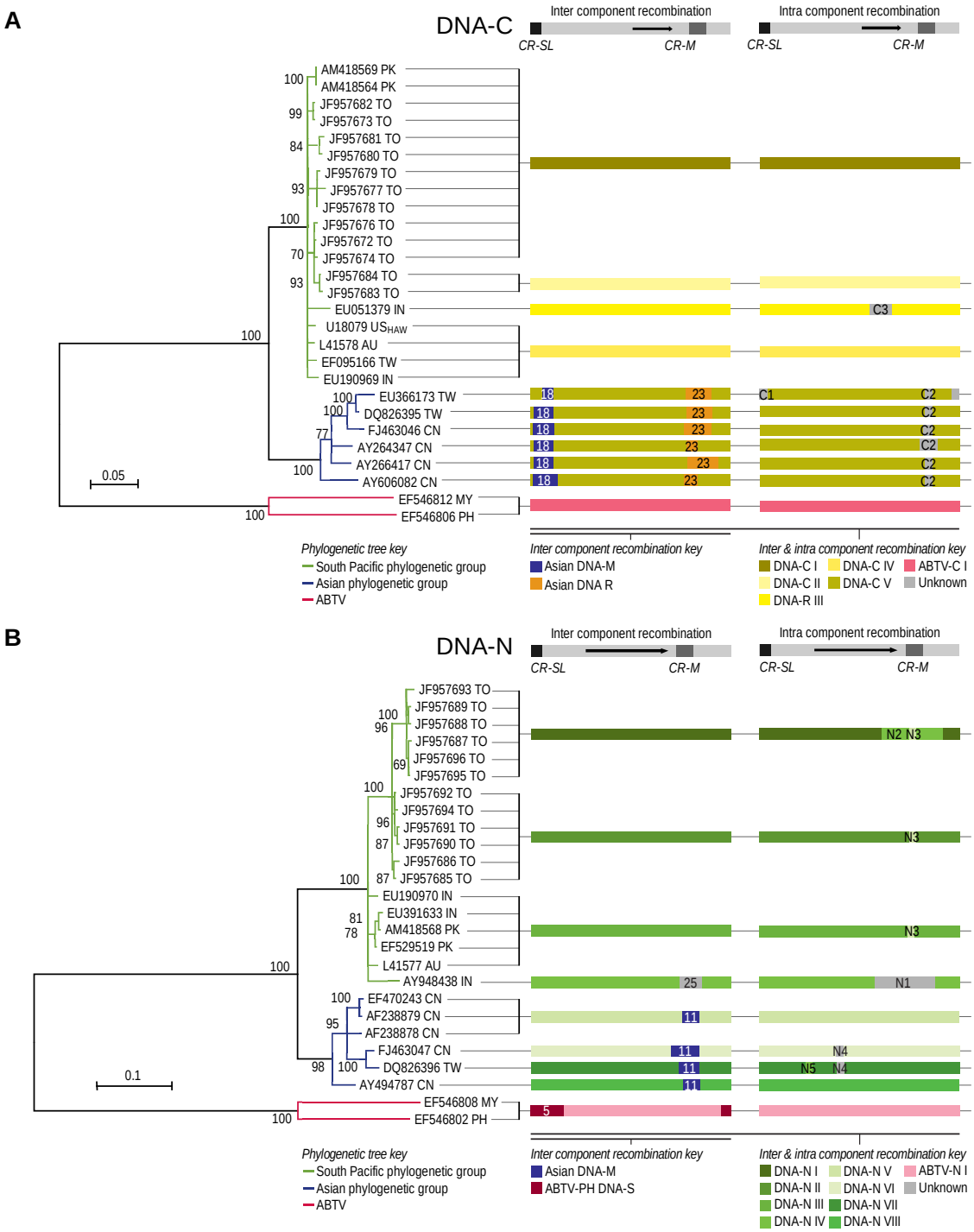


Figure 2.6: A. Midpoint rooted maximum likelihood phylogenetic tree (nucleotide substitution model GTR+G4) of BBTV DNA-C components. B. Midpoint rooted maximum likelihood phylogenetic tree (nucleotide substitution model GTR+G4) of BBTV DNA-N components. Bootstrap support (1000 bootstrap replicates) is shown for each branch with >60% support. The South Pacific and Asian phylogenetic group sequences are indicated by green and blue branches, respectively. ABTV isolates are indicated in red. Two letter international country codes are provided with GenBank accession numbers. Sequences identified as either inter- or intra-component recombinants by RDP4 are shown with their corresponding breakpoint positions. Details of recombination break points and statistical support for evidence of recombination are provided in Supplementary Tables 2.1 and 2.2. Note that the intra-component recombination events labelled C2 and N1 (with "unknown" parental sequences) are likely respectively conflated with the inter-component recombination events labelled 23 and 25.

2.4.2 Reassortment

Each BBTV genome component is packaged into a separate virion, and the probable transmission of large numbers of virions provides an ideal environment for the reassortment of genome components. Reassortment has previously been shown in multicomponent plant RNA viruses (Chen *et al.*, 2007; Gu *et al.*, 2007; Maoka *et al.*, 2010) and other DNA viruses (Chen *et al.*, 2009; Pita *et al.*, 2001; Saunders *et al.*, 2002) and evidence suggests that the reassortment of genome components may occur during BBTV evolution (Bell *et al.*, 2002); (Horser *et al.*, 2001b); (Hu *et al.*, 2007). Importantly, reassortment has been implicated in host adaption and the evolution of increased virulence in various other plant viruses. For example, reassortment between two *Tomato spotted wilt virus* (TSWV, a tripartite RNA virus) strains has been observed to result in a strain capable of breaking the TSWV resistance of some tomato cultivars (Qiu & Moyer, 1999). Therefore reassortment may play an important adaptive role during BBTV evolution.

We investigated evidence of reassortment in BBTV isolates by analysing concatenated genome components using the recombination detection program, RDP4 (Martin *et al.*, 2010) (Table 2.1). We found well-supported evidence of eight reassortment events amongst the analysed BBTV isolates (Figure 2.7). Reassortment events can be seen within and across the SPG and AG. All but one of these events has involved the reassortment of a single component. The exception, event five in Figure 2.7, appears to have involved both the DNA-R and DNA-N components.

In the case of isolate V-1 (24-TW) from Taiwan, we also found evidence of two reassortment events (see events one and two in Figure 2.7) but with each event apparently having involved different parental genomes. One event involved the acquisition of a DNA-M component from an AG–like parental genome. The origin of the genome component acquired during the other event is, however, less easily traceable as two versions of DNA-R, both obtained from the same infection, have been deposited for this isolate (Hu *et al.*, 2007). Therefore our interpretation of this event is dependent on which of the specific component clones were used to generate the analysed concatemers. For example, DNA-R clone V-1 (EF095161), like its DNA-M counterpart is derived from an AG–like parental genome, whereas V-1a (EF095162) is derived from SPG-like parental genome. This evidence is also clear from the ML phylogenetic trees of these components (Figures 2.2-2.6). Hu *et al.* (2007) have also

suggested that, on the basis of their phylogentic analyses, this isolate is a reassortant (Hu *et al.*, 2007).

In the case of the Australian Nambour genome (4-AU) we found that all of its components other than DNA-M had most likely been derived from viruses resembling the Thatha (1-PK), Lucknow (3-IN) and Robusta BT-1 (22-IN) isolates whereas its DNA-M component has likely been derived from a virus resembling the Tongan isolates (see event three in Figure 2.7; Figure 2.5). The DNA-N component of the Lucknow (3-IN) isolate has likely been derived from a SPG genome resembling those of the Thatha (1-PK) and Nambour (4-AU) isolates (see event four in Figure 2.7). Similarly, the Haikou (23-CN) isolate has apparently acquired its DNA-S component from another isolate within the AG resembling NSP (19-CN; see event seven in Figure 2.7).

Within the Tongan sequences we found evidence of three probable reassortment events. The Tongan TO208-Vava'u isolate (44-TO; see event five in Figure 2.7) has DNA-R and -N components that are likely derived from an isolate (or isolates) resembling those which we sampled on the island of Tongatapu. This is clear evidence of inter-island BBTV transfers leading to component reassortments. Although no minor parent was able to be identified in event six, there is evidence that the common ancestral sequence of the Tongan isolates (from Tongatapu and an isolate from Ha'apai) have acquired a DNA-M component from an isolate distantly resembling viruses that have today been found on the island of Vava'u, in India, in Pakistan and in Egypt (Figure 2.7).

Interestingly, the Australian Nambour isolate has a DNA-M component that has apparently been derived from a genome resembling those which we have characterised from Tongatapu (Figure 2.5). Finally, the DNA-U3 component of the Vava'u island isolates (44-TO, 45-TO and 46-TO) is likely derived from a genome resembling that of the Australian Nambour isolate (see event eight in Figure 2.7). Our ability to detect all of these BBTV genome reassortments despite the few full genomes available for analysis, strongly suggests that reassortment may play some crucial role in the adaptive evolution of BBTV genomes.

Our use of standard recombination detection tools to analyse concatemerised babuvirus genome components for evidence of reassortment provides an important validation of similar results obtained previously by Hu *et al.* (2007) using phylogenetics-based reassortment detection approaches. Whereas evidence of reassortment can certainly be detected by

comparing phylogenetic trees constructed from different genome components, naturally occurring homologous recombination between the homologous components of different virus species can cause alterations in tree topologies that could easily be misidentified as reassortments. Also, basic phylogenetic tree comparisons cannot yield any estimates of statistical support for whether recombination has occurred or not (although more sophisticated Bayesian inference approaches may be able to achieve this) because of multiple testing issues. The method applied here both accounts for homologous recombination between components and provides strong statistical evidence of reassortment.

Event	Component						Reassortment Sequence(s)	Sequence(s) used to infer minor Parent(s)	Sequence(s) used to infer major parent(s)	Methods	p-value
	R	U3	S	M	C	N					
1	o	o	o	O	o	-	24-TW	5-TW 21-CN 23-CN 25-TW	All Tongan 1-PK 3-IN 4-AU 22-IN 33-IN	R**G**BMCST	$2.52x10^{-52}$
2	O	o	o	o	o	-	24-TW	5-TW 19-CN 21-CN 23-CN 25-TW	Unknown	R**G**BMCST	2.048×10^{-50}
3	o	o	o	O	o	o	4-AU	Unknown	1-PK 3-IN 22-IN	RG**B**MCST	$9.15x10^{-15}$
4	o	o	o	o	o	O	3-IN	1-PK 4-AU	All Tongan	RGBMC**S**T	$2.13x10^{-16}$
5	O	o	o	o	o	O	44-TO	35-TO 36-TO 37-TO 38-TO 40-TO 41-TO	1-PK 4-AU 22-IN 33-IN	RGBMC**S**T	$1.61x10^{-16}$
6	o	o	o	O	o	o	35-TO 36-TO 37-TO 38-TO 39-TO 40-TO 41-TO	Unknown	42-TO 43-TO 44-TO 45-TO 46-TO	RGBMCS**T**	$5.68x10^{-16}$
7	o	o	O	o	o	o	23-CN	19-CN	20-CN	MC**S**	$2.11x10^{-08}$
8	o	O	o	o	o	o	44-TO 45-TO 46-TO	4-AU	35-TO 36-TO 38-TO 39-TO 40-TO 41-TO 42-TO 43-TO	RBMCS**T**	$1.68x10^{-05}$

Methods used by RDP4 to infer reassortment as denoted by R, G, B, M, C, S and T indicate recombination detection by the RDP, GENCONV, BOOTSCAN, MAXCHI, CHIMERA, SISCAN and 3SEQ methods, respectively and the p-value shown is for the method indicated in bold.

O non-reassortment sequence (South Pacific group)

O reassortment sequence (South Pacific group)

O non-reassortment sequence (Asian group)

O reassortment sequence (Asian group)

Figure 2.7: Reassortment analyses. Circles depict sequences present within the reassortant genomes. Large circles are derived from a minor parent, and small, circles from the major parent.

2.4.3 Inter-component recombination (recombination between non-homologous genome components)

It has been found previously in both multi-component geminiviruses (Saunders *et al*., 2002; Hou and Gilbertson, 1996; Roberts and Stanley, 1994; Jovel *et al*., 2004) and in the nanoviruses (Hu *et al*., 2007; Hughes, 2004; Hyder *et al*., 2011) that inter-component recombination is frequent and is probably a vital mechanism via which Rep containing components retain trans-replicational control over their helper components. An overview of the breakpoint distributions revealed by our inter-component recombination analysis of BBTV genome components (results of which are summarised in Figures 2.2-2.6 and in Supplementary Table 2.1) indicated the presence of a major recombination hot-spot around the CR-M region and a minor hotspot around the CR-SL (Figure 2.1B). Overall, we observed a greater degree of inter-component recombination in components DNA-U3 and DNA-M but particularly interesting examples of inter-component recombinants in other components include (1) the DNA-R of isolate TW3 (EU366169; event 13 in Figure 2.2) which has a fragment of a DNA-M-like sequence in the domain spanning the CR-M and CR-SL; and (2) the DNA-R of the ABTV isolate from the Philippines has a region spanning the CR-M that is most similar to the DNA-N sequence of the Malaysian ABTV isolate (event 19 in Figure 2.2).

2.4.4 Intra-component recombination (recombination between homologous genome components)

Although recombination amongst homologous nanovirus and babuvirus genome components has been detected in various other studies (Hu *et al*., 2007; Hyder *et al*., 2011; Timchenko *et al*., 2000) we sought to reanalyse recombination in BBTV taking into account the fact that many of the recombination signals detected in these other studies likely originated through inter- rather than intra-component recombination. All the SPG DNA-R components appear to have descended from a common recombinant ancestor that had derived a portion of its Rep encoding sequence from an unknown babuvirus (the event labelled R2 in Figure 2.2 and Supplementary Table 2.2). Similarly, the DNA-R components of the AG isolates, bp26 (AB250955; event R1 in Figure 2.2), IG64 (AB186924) and IJs11 (AB186926; the latter two carry evidence of the same event labelled R4 in Figure 2.2) contain small fragments of sequence apparently derived from parental viruses in the SPG. Importantly, our inter-component recombination analyses clearly indicated that the recombination event detected in

the Taiwanese isolate, TW3 (EU366169; labelled R3 in Figure 2.2) is very likely the same inter-component recombination event labelled as event 13 in our inter-component analysis. Wherever such conflation of inter- and intra-component recombination signals occurred in our analyses (these are all pointed out in the legends of Figures 2.2 through 2.6), it is very likely that these signals are the product of inter-component recombination events being interpreted by the recombination analysis software as being fragments of sequence derived from a highly divergent parental sequence.

The most recombinogenic of the BBTV components seems to be U3 within which we found evidence of at least six genuine intra-component recombination events (events U1, U6, U7, U8, U9 and U10 in Figure 2.3). Previous analyses of recombination amongst DNA-U3 sequences (Hyder *et al.*, 2011) have similarly identified frequent recombination in this genome component. Crucially, however, our analysis of multiple genome components within the same alignment indicated that many apparent intra-component recombination events within the CR-M and CR-SL were in fact inter-component recombination events (these are labelled U2, U3, U4 and U5 in Figure 2.3).

After accounting for likely inter-component recombination events we found evidence of two intra-component recombination events amongst the BBTV DNA-S sequences (Figure 2.4), four amongst the BBTV DNA-M sequences (Figure 2.5), one amongst the BBTV DNA-C sequences (Figure 2.6A) and four amongst the BBTV DNA-N sequences (Figure 2.6B). Most notable amongst the detected recombination events are: (1) an event involving the transfer of a DNA-S fragment (labelled S3 in Figure 2.4) from an AG isolate into a SPG isolate to yield a recombinant that is apparently widely circulating in Pakistan; (2) an event involving the transfer of a small DNA-N fragment from an AG-like ancestral sequence into the progenitor of all the SPG DNA-N sequences (event N3 in Figure 2.6B), (3) an event involving the transfer of a DNA-N fragment from a SPG virus resembling those found in India, Pakistan and Australia into the progenitor of a group of viruses from Tonga and (4) an event involving the cross-species transfer of a large apparently AG BBTV derived DNA-M fragment (labelled 24 in Figure 2.5) into the DNA-M of ABTV.

Since both inter- and intra-component recombination events are readily detectable in the small number of available BBTV component sequences that we have analysed, it is likely that, along with component reassortment, recombination is a major evolutionary process driving the diversification of BBTV. Although there are still too few recombination events

detectable within the available BBTV sequences to be sure, our results indicate that either recombination frequencies or the selective benefits of recombination might vary from component to component. Components encoding core functions, such as replication proteins (DNA-R), seem to be more prone to intra- rather than inter-component recombination. On the other hand, DNA-U3 exhibits complex intra- and inter-recombination patterns. These observations suggest that recombination in components such as DNA-U3 may be selectively more favourable (or rather selectively less deleterious) than recombination in components such as DNA-R and DNA-C.

In order for recombination to occur, multiple variants must be present within the same cell during replication. Recombination of nanoviruses could occur either by strand displacement of replication complexes and reattachment to different template molecules during rolling circle replication or by the rescue of incompletely replicated or fragmented genomes via host mediated double-stranded break repair pathways (Jeske *et al.*, 2001). Host-specific selection acting on these recombined sequences could then hasten either their fixation (positive selection) or removal (negative selection) from the population (Martin *et al.*, 2011). The fact that many of the recombination events which we have characterized (whether intra- or inter-component recombinants) are detectable within the genomes of multiple circulating viruses clearly indicates that at least some of the recombination events we have detected may have been adaptive enough to have been selectively favoured. Amplification and sequencing of a larger number of components from individual plants dually infected with two or more BBTV strains should provide some insights into the extent of inter- and intra-component recombination arising during individual mixed infections. However, the nature of the exponential amplification tools that are widely employed for cloning ssDNA viruses would mean sampling of only the potentially minor subset of recombinants that become major population members within such infections. Ideally deep sequencing using next-generation sequencing methods may prove to be a better option to analyse such recombinants.

2.5 Concluding remarks

Twelve full genomes have been sequenced from BBTV infected banana plants from the Kingdom of Tonga, all of which fall within the SPG. Using various statistical tests for recombination to analyse these Tongan genomes, together with those available in GenBank, we found evidence of frequent genome component reassortment and both intra- and inter-component homologous recombination events. Our study highlights the benefits of characterizing complete BBTV genomes rather than focusing on individual components. The clear evidence we have found of reassortment and recombination events within and between the AG and SPG support the hypothesis that both groups may share a common geographic origin and/or have present geographical ranges that are largely overlapping. Whereas mixed infections involving babuviruses from different groups or even different species must be relatively common to have permitted the patterns of sequence and component transfers that we have observed, both variations in the recombination and reassortment patterns seen for different components and the occurrence of circulating recombinant/reassortant forms suggest genome sequence transfers in babuviruses might have substantial fitness effects.

GenBank Accession numbers: JF957625-JF957696

Supplementary table 2.1: Inter-component recombination analysis, R, G, B, M, C, S, T indicate recombination detection by the RDP, GENCONV, BOOTSCAN, MAXCHI, CHIMERA, SISCAN and 3SEQ methods, respectively and the p-value shown is for the method indicated in bold. # Trace evidence was identified for this sequence. [P] evidence of partial recombination. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence.

Event	Breakpoint	Recombinant Sequence(s)	Sequence(s) used to infer Major Parent(s)	Sequence(s) used to infer Minor Parent(s)	Detection methods	p-value
1	785-1047	M-JF957668-TO	Unknown	All South Pacific DNA S	R**G**BMCST	$2.41x10^{-41}$
2	790-1045	M-AY494788-CN	Unknown	All Asian DNA S	R**G**BMCST	$8.16x10^{-36}$
3	926-75	U3-JF957637-TO	All South Pacific DNA-U3 except recombinants	All Asian DNA C except *AY264347, AY266417*	RGBMCS**T**	$4.22x10^{-29}$
		U3-JF957640-TO	U3-FJ463043-TW	All TO DNA C except *JF957682, JF657681*		
		U3-JF957648-TO	U3-AY606084-CN	All PK DNA C		
				C-EF095166-TW		
4	20-161	S-EF546810-abaca-767	All South Pacific DNA S	N-EF546808-abaca-767	R**G**BMCST	$9.88x10^{-24}$
			All Asian DNA S			
5	1045-157	N-EF546802-abaca-1108	All South Pacific DNA N	S-EF546804-abaca-1108	RG**B**MCST	$5.29x10^{-27}$
6	883-216	U3-EU366170-TW	U3-FJ773283-TW	All South Pacific DNA N	RGBMCS**T**	$1.62x10^{-42}$
			U3-FJ009239-IN	All Asian DNA N		
			U3-AY996563-PK	N-EF546808-abaca-767		
			U3-DQ826392-TW			
			U3-FJ463043-TW			
			U3-EU140341-IN			
			U3-AY960129-IN			
			U3-EF095163-TW			
			U3-JF957643-TO			
			U3-JF957642-TO			
			U3-JF957641-TO			
			U3-U18078-US-HAW			
			U3-JF957644-TO			
			U3-JF957646-TO			
			U3-JF957645-TO			
			U3-JF957647-TO			
			U3-JF957639-TO			
			U3-JF957638-TO			
			U3-AY606084-CN			
7	750-893	S-AY494786-CN	All South Pacific DNA S	All CN DNA N	**R**GBMCST	$4.60x10^{-16}$
		M-AY494788-CN[P]		N-DQ826396-TW		
		S-AF148945-VN[P]		N-L41577-AU		
		S-AB113661-VN[P]				
		S-AB113662-VN				
		S-AF238877-CN				
		S-AF238876-CN				
		S-FJ463044-CN				
		S-EU366171-TW				
		S-AB186929-ID				
		S-AB186927-ID				
		S-AB186928-ID				
		S-AF148942-TW				
		S-DQ826393-TW				
		S-AB108451-JP				
		S-AB108449-JP				
		S-AB108450-JP				
		S-AF148068-PH				
		S-AB250958-PH				
		S-AB250957-PH				
		S-AB250956-PH				
		S-AB189068-PH				
8	802-1038	All South Pacific DNA S	Unknown	All South Pacific DNA M except *JF957668*	R**G**BMCST	$1.60x10^{-30}$

9	1021*-67	U3-FJ009239-IN U3-AY996563-PK U3-FJ773283-TW U3-DQ826392-TW U3-FJ463043-TW U3-DQ826391-TW[T]	U3-EU046323-IN U3-EU402601-IN U3-EU140341-IN U3-AY960129-IN U3-EF095163-TW U3-L41576-AU U3-JF957643-TO U3-JF957642-TO U3-JF957641-TO U3-U18078-US-HAW U3-JF957644-TO U3-JF957646-TO U3-JF957645-TO U3-JF957647-TO U3-JF957639-TO U3-JF957638-TO U3-AY606084-CN	All South Pacific DNA S All Asian DNA S S-EF546804-abaca-1108 R-EU366169-TW R-AB252641-MM M-AF349568-CN M-JF957672-TO M-JF957671-TO M-L41575-AU M-JF957661-TO M-JF957662-TO M-JF957664-TO M-JF957663-TO M-JF957665-TO M-JF957667-TO	RG**B**	$6.50x10^{-15}$
10	750-1037	U3-AY606084-CN U3-FJ773283-TW U3-DQ826392-TW U3-FJ463043-TW U3-DQ826391-TW[T] U3-EU366170-TW[P]	Unknown	All Asian DNA R	**R**GBMCST	$2.17x10^{-13}$
11	828-913	N-AY494787-CN N-FJ463047-CN N-DQ826396-TW N-AF238878-CN N-EF470243-CN N-AF238879-CN	All South Pacific DNA N	All TW DNA M and M-FJ463045-CN M-AF349568-CN	R**G**BMS	$2.17x10^{-11}$
12	974-1033	M-EF546811-abaca-767	Unknown	R-EF546813-abaca-767	R**G**BMCST	$4.06x10^{-12}$
13	1069*-49	R-EU366169-TW	All South Pacific DNA R R-AF238874-CN R-AF246123-CN R-AF416477-VN R-AB113659-VN R-U97525-CN R-EF546813-abaca-767	M-EU366172-TW M-JF957672-TO M-JF957671-TO M-L41575-AU M-JF957661-TO M-JF957662-TO M-JF957664-TO M-JF957663-TO M-JF957665-TO M-AF102783-EG M-JF957667-TO	RG**B**	$7.31x10^{-10}$
14	953-8	M-EF546805-abaca-1108	Unknown	R-EF546807-abaca-1108	**R**GB	$5.12x10^{-09}$
15	748*-806	U3-JF957647-TO U3-JF957640-TO U3-JF957648-TO U3-JF957637-TO U3-U18078-US-HAW	U3-AY606084-CN U3-DQ826392-TW U3-FJ463043-TW	M-JF957668-TO N-L41577-AU N-EU190970-IN N-EU391633-IN N-EF529519-PK N-JF957694-TO N-JF957692-TO N-JF957691-TO N-JF957690-TO N-JF957686-TO N-JF957687-TO N-JF957696-TO N-JF957695-TO N-JF957689-TO N-JF957688-TO All South Pacific DNA S except *L41574-AU, AM418540-PK*	**G**ST	$3.59x10^{-05}$
16	877*-1040	M-FJ463045-CN M-JF957672-TO M-JF957671-TO M-L41575-AU	Unknown	All Asian DNA S	RGBMC**S**	$2.15x10^{-25}$

		M-JF957661-TO M-JF957662-TO M-JF957664-TO M-JF957663-TO M-JF957665-TO M-AF102783-EG M-AM418541-PK M-EU190971-IN M-AY953429-IN M-JF957670-TO M-JF957669-TO M-JF957667-TO M-JF957666-TO M-EU516323-IN M-AY948439-IN M-AF349568-CN M-EU366172-TW M-EF095165-TW M-DQ826394-TW				
17	798-846*	M-AM418541-PK M-JF957672-TO M-JF957671-TO M-L41575-AU M-JF957661-TO M-JF957662-TO M-JF957664-TO M-JF957663-TO M-JF957665-TO M-AF102783-EG M-EU190971-IN M-AY953429-IN M-JF957670-TO M-JF957669-TO M-JF957667-TO M-JF957666-TO M-EU516323-IN M-AY948439-IN	Unknown	All South Pacific DNA R	R**G**BM	$7.63x10^{-13}$
18	24*-86	C-DQ826395-TW C-AY606082-CN C-FJ463046-CN C-EU366173-TW C-AY266417-CN C-AY264347-CN[T]	Unknown (M)	All CN DNA M M-DQ826394-TW M-EF095165-TW	R**G**S	$6.44x10^{-11}$
19	1001-1063	R-EF546807-abaca-1108	R-EF546813-abaca-767	N-EF546808-abaca-767	R**B**MT	$1.44x10^{-09}$
20	772-813	U3-JF957639-TO U3-JF957646-TO U3-JF957645-TO U3-JF957638-TO	Unknown	All South Pacific DNA R	R**G**BM	$1.36x10^{-10}$
21	794-868*	M-AF349568-CN M-FJ463045-CN M-EU366172-TW M-EF095165-TW M-DQ826394-TW	Unknown	R-AF416473-VN R-AY450396-CN R-AF416472-VN	**R**GBM	$6.90x10^{-09}$
22	1031*-106	M-EU366172-TW	M-EF095165-TW M-AY494788-CN M-FJ463045-CN	Unknown	RGB**T**	$2.31x10^{-10}$
23	852-984	C-FJ463046-CN C-AY606082-CN[T] C-EU366173-TW C-DQ826395-TW C-AY266417-CN C-AY264347-CN[T]	Unknown	All Asian DNA R except *EU366169*	R**G**B	$2.48x10^{-08}$
24	62*-391	M-EF546811-abaca-767	All BBTV DNA R except	M-AF349568-CN	MC**S**	$7.79x10^{-12}$

			AB252641-MM			
			All Abaca DNA R	M-AY948439-IN M-FJ463045-CN M-EU366172-TW M-DQ826394-TW		
25	811-922	N-AY948438-IN	N-EU190970-IN N-L41577-AU N-EU391633-IN N-EF529519-PK N-AM418568-PK	Unknown	R**G**BMST	$1.52x10^{-07}$
26	774-815	U3-FJ009239-IN U3-AY996563-PK U3-EU402601-IN U3-EU140341-IN U3-EU046323-IN U3-AY960129-IN U3-EF095163-TW U3-L41576-AU U3-JF957643-TO U3-JF957642-TO[T] U3-JF957641-TO U3-JF957644-TO	Unknown	R-AF416474-VN All South Pacific DNA R except *JF957627-TO* *JF957628-TO* *AF416467-TO*	R**G**B	$3.46x10^{-07}$

Supplementary table 2.2: Intra-component recombination analysis. R, G, B, M, C, S, T indicate recombination detection by the RDP, GENCONV, BOOTSCAN, MAXCHI, CHIMERA, SISCAN and 3SEQ methods, respectively and the p-value shown is for the method indicated in bold. # Trace evidence was identified for this sequence. [P] evidence of partial recombination. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence.

Event	Breakpoint	Recombinant Sequence(s)	Sequence(s) used to infer Major Parent(s)	Sequence(s) used to infer Minor Parent(s)	Detection methods	p-value
DNA-R 1	1050-18	R-AB250955-PH	R-EU366169-TW	R-AF416467-TO	**R**GT	$2.61x10^{-04}$
				R-AB252640-MM		
				R-AF102780-EG		
				R-AF416465-EG		
				R-AF416466-FJ		
				R-AF416470-IN		
				R-AM418534-PK		
				R-AM418535-PK		
				R-AM418536-PK		
				R-AM418537-PK		
				R-AM418538-PK		
				R-AM418539-PK		
				R-AY996562-PK		
				R-DQ656118-IN		
				R-DQ656119-IN		
				R-EF095162-TW		
				R-EU140342-IN		
				R-FJ009238-IN		
				R-FJ009240-IN		
				R-JF957627-TO		
				R-JF957628-TO		
				R-JF957629-TO		
				R-JF957630-TO		
				R-JF957631-TO		
				R-JF957632-TO		
				R-JF957634-TO		
				R-JF957635-TO		
				R-JF957636-TO		
				R-S56276-AU		
				R-U18077-US-HAW		
DNA-R 2	718-1057	R-DQ256267-IN	R-AF110266-CN	Unknown	M**S**	$8.27x10^{-08}$
		R-AB252639-MM	R-AB108452-JP			
		R-AB252640-MM[T]	R-AB108453-JP			
		R-AB252641-MM	R-AB108455-JP			
		R-AF416465-EG	R-AB108456-JP			
		R-AF416466-FJ	R-AB108457-JP			
		R-AF416467-TO	R-AB108458-JP			
		R-AF416470-IN	R-AB113659-VN			
		R-AM418534-PK[T]	R-AB113660-VN			
		R-AM418535-PK[T]	R-AB186924-ID			
		R-AM418536-PK[T]	R-AB189067-PH			
		R-AM418537-PK[T]	R-AB250953-PH			
		R-AM418538-PK	R-AB250954-PH			
		R-AM418539-PK[T]	R-AB250955-PH			
		R-AY996562-PK[T]	R-AF238874-CN			
		R-DQ656118-IN[T]	R-AF238875-CN			
		R-DQ656119-IN[T]	R-AF246123-CN			
		R-EF095162-TW	R-AF416464-VN			
		R-EU140342-IN	R-AF416468-TW			
		R-FJ009238-IN[T]	R-AF416469-PH			
		R-FJ009240-IN[T]	R-AF416479-VN			
		R-JF957625-TO	R-DQ826390-TW			
		R-JF957626-TO[T]	R-EF095161-TW			

		R-JF957627-TO[T] R-JF957628-TO[T] R-JF957629-TO[T] R-JF957630-TO[T] R-JF957631-TO R-JF957632-TO[T] R-JF957633-TO[T] R-JF957634-TO R-JF957635-TO[T] R-JF957636-TO[T] R-S56276-AU[T] R-U18077-US-HAW	R-EU366169-TW R-FJ463042-CN			
DNA-R 3	1002-228	R-EU366169-TW	R-AB186924-ID R-AB108453-JP R-AB108455-JP R-AB108456-JP R-AB108457-JP R-AB189067-PH R-AB250953-PH R-AB250954-PH	Unknown	**M**S	$1.96x10^{-03}$
DNA-R 4	206-303	R-AB186925-ID R-AB186926-ID	Unknown	R-JF957631-TO R-U18077-US-HAW	R**T**	$1.18x10^{-02}$
DNA-U3 1	940-4	U3-AY606084-CN	U3-FJ463043-TW U3-DQ826392-TW U3-DQ826391-TW U3-FJ773283-TW	U3-EF095163-TW U3-AY960129-IN U3-EU402601-IN U3-JF957638-TO U3-JF957639-TO U3-JF957647-TO U3-L41576-AU U3-U18078-US-HAW U3-JF957646-TO U3-EU140341-IN U3-EU046323-IN U3-JF957641-TO U3-JF957642-TO U3-JF957643-TO U3-JF957644-TO U3-JF957645-TO	R**G**BMCST	$4.07x10^{-16}$
DNA-U3 2	1012-4	U3-FJ009239-IN U3-AY996563-PK	U3-EU140341-IN U3-AY960129-IN U3-EF095163-TW U3-EU402601-IN U3-JF957638-TO U3-JF957639-TO U3-JF957647-TO U3-L41576-AU U3-U18078-US-HAW U3-JF957646-TO U3-EU046323-IN U3-JF957641-TO U3-JF957642-TO U3-JF957643-TO U3-JF957644-TO U3-JF957645-TO	U3-DQ826392-TW U3-FJ463043-TW U3-FJ773283-TW	RG**B**MCT	$6.87x10^{-15}$
DNA-U3 3	949-1026	U3-DQ826391-TW	U3-DQ826392-TW U3-FJ463043-TW	U3-JF957639-TO U3-AY960129-IN U3-EF095163-TW U3-EU402601-IN U3-JF957638-TO U3-JF957647-TO U3-L41576-AU U3-U18078-US-HAW U3-AY996563-PK U3-FJ009239-IN	R**M**T	$3.64x10^{-08}$

				U3-JF957646-TO U3-EU140341-IN U3-EU046323-IN U3-JF957641-TO U3-JF957642-TO U3-JF957643-TO U3-JF957644-TO U3-JF957645-TO U3-EF546809-abaca-767		
DNA-U3 4	911*-11	U3-FJ463043-TW U3-DQ826392-TW[T] U3-FJ773283-TW[T]	U3-JF957639-TO U3-AY960129-IN U3-EF095163-TW U3-JF957638-TO U3-JF957647-TO U3-L41576-AU U3-U18078-US-HAW U3-AY996563-PK U3-JF957646-TO U3-EU140341-IN U3-EU046323-IN U3-JF957641-TO U3-JF957642-TO U3-JF957643-TO U3-JF957644-TO U3-JF957645-TO	Unknown	R**M**CT	$6.53x10^{-07}$
DNA-U3 5	944-14	U3-JF957640-TO U3-JF957637-TO U3-JF957648-TO	U3-JF957647-TO U3-EF095163-TW U3-JF957638-TO U3-JF957639-TO U3-L41576-AU U3-U18078-US-HAW U3-JF957646-TO U3-JF957642-TO U3-JF957644-TO U3-JF957645-TO	Unknown	RGMC**S**T	$2.29x10^{-17}$
DNA-U3 6	692-751	U3-AY606084-CN	Unknown	U3-DQ826391-TW U3-DQ826392-TW	R**G**T	$1.95x10^{-05}$
DNA-U3 7	792-815	U3-FJ773283-TW U3-DQ826392-TW U3-FJ463043-TW U3-DQ826391-TW[T]	U3-U18078-US-HAW U3-AY960129-IN U3-EF095163-TW U3-EU402601-IN U3-JF957638-TO U3-JF957639-TO U3-JF957647-TO U3-L41576-AU U3-AY996563-PK U3-FJ009239-IN U3-JF957646-TO U3-EU140341-IN U3-EU046323-IN U3-JF957641-TO U3-JF957642-TO U3-JF957643-TO U3-JF957644-TO U3-JF957645-TO U3-JF957637-TO U3-JF957640-TO U3-JF957648-TO	U3-EF546803-abaca-Q1108	**R**	$3.27x10^{-03}$
DNA-U3 8	998-568	U3-EU140341-IN	U3-JF957642-TO	Unknown	MC**S**	$1.27x10^{-04}$
DNA-U3 9	301*-340*	U3-EF546803-abaca-Q1108	U3-EF546809-abaca-767	Unknown	RG**M**	$8.60x10^{-03}$
DNA-U3 10	612-760	U3-EU046323-IN	U3-AY960129-IN	Unknown	**R**GBS	$7.60x10^{-03}$
DNA-U3 11	571-867*	U3-FJ009239-IN U3-AY960129-IN	U3-EU402601-IN	U3-L41576-AU U3-JF957643-TO	RGMC**S**	$1.52x10^{-09}$

		U3-AY996563-PK				
DNA-S 1	983-418	S-AF148945-VN S-AB113661-VN S-AB113662-VN	S-AF238877-CN S-AB108450-JP S-AB189068-PH S-AB250956-PH S-AF148068-PH S-DQ826393-TW S-EU366171-TW	Unknown	**M**C	$4.69x10^{-05}$
DNA-S 2	34-243	S-EF546804-abaca-1108	S-EF546810-abaca-767	S-AB252644-MM S-AB108449-JP S-AB108450-JP S-AB108451-JP S-AB186927-ID S-AB186928-ID S-AB186929-ID S-AB189068-PH S-AB250956-PH S-AB250957-PH S-AB250958-PH S-AB252642-MM S-AB252643-MM S-AF148068-PH S-AF148942-TW S-AF148943-BI S-AF148944-FJ S-AF238876-CN S-AF238877-CN S-AM418540-PK S-AM418565-PK S-AM418566-PK S-AM418567-PK S-AY494786-CN S-DQ826393-TW S-EF095164-TW S-EF687856-IN S-EU366171-TW S-FJ463044-CN S-GQ249344-CM S-JF957673-TO S-JF957674-TO S-JF957675-TO S-JF957676-TO S-JF957677-TO S-JF957678-TO S-JF957679-TO S-JF957680-TO S-JF957681-TO S-JF957682-TO S-JF957683-TO S-JF957684-TO S-L41574-AU	M**S**	$2.34x10^{-26}$
DNA-S 3	294-420	S-AM418540-PK S-AM418565-PK S-AM418566-PK S-AM418567-PK S-GQ249344-CM[T]	S-JF957677-TO S-AB252644-MM S-AF148943-BI S-EF687856-IN S-JF957673-TO S-JF957675-TO S-JF957678-TO S-JF957680-TO S-L41574-AU	S-AB108451-JP S-AB108449-JP S-AB108450-JP S-AB186927-ID S-AB186928-ID S-AB186929-ID S-AB189068-PH S-AB250956-PH S-AB250957-PH S-AB250958-PH S-AF148068-PH S-AF148942-TW	RG**M**	$7.74x10^{-03}$
DNA-M 1	278-373	M-JF957665-TO	M-JF957666-TO	Unknown	**R**GMCT	$4.59x10^{-09}$

			M-L41575-AU M-JF957672-TO M-JF957671-TO M-JF957670-TO M-JF957669-TO M-JF957668-TO M-JF957667-TO M-JF957664-TO M-JF957663-TO M-JF957662-TO M-JF957661-TO M-EU516323-IN M-EU190971-IN M-AY953429-IN M-AY948439-IN M-AM418541-PK M-AF102783-EG			
DNA-M 2	757-47	M-AF102783-EG	M-AM418541-PK M-JF957670-TO M-EU516323-IN M-AY953429-IN M-AY948439-IN	M-JF957668-TO M-L41575-AU M-JF957672-TO M-JF957671-TO M-JF957664-TO M-JF957663-TO M-JF957661-TO	**M**T	$3.09x10^{-04}$
DNA-M 3	987-82	M-EU366172-TW	M-FJ463045-CN M-EF095165-TW	Unknown	**G**	$4.91x10^{-03}$
DNA-M 4	462-652	M-AY494788-CN	M-AF349568-CN M-EU366172-TW M-DQ826394-TW	Unknown	**M**	$1.27x10^{-02}$
DNA-M 5	990-8	M-DQ826394-TW	M-EF095165-TW M-JF957667-TO M-JF957666-TO M-JF957665-TO M-FJ463045-CN M-AY494788-CN M-AF349568-CN	Unknown	R**G**	$8.34x10^{-08}$
DNA-M 6	318-418	M-EF546811-abaca-767	M-EF546805-abaca-1108	Unknown	**M**	$1.67x10^{-02}$
DNA-M 7	861-905	M-EF546811-abaca-767	M-EF546805-abaca-1108	Unknown	**M**	$1.60x10^{-02}$
DNA-M 8	1006-13*	M-EF546811-abaca-767	M-EF546805-abaca-1108	M-AF349568-CN M-FJ463045-CN M-EF095165-TW	**R**	$1.60x10^{-02}$
DNA-C 1	958-28	C-EU366173-TW	C-AY264347-CN C-AY266417-CN C-FJ463046-CN	Unknown	R**G**	$5.68x10^{-05}$
DNA-C 2	822-855	C-DQ826395-TW C-AY264347-CN C-AY266417-CN C-AY606082-CN[T] C-EU366173-TW C-FJ463046-CN[T]	C-JF957679-TO C-JF957672-TO C-JF957673-TO C-JF957674-TO C-JF957676-TO C-JF957678-TO C-JF957683-TO C-JF957684-TO C-U18079-US-HAW	Unknown	**G**M	$5.14x10^{-04}$
DNA-C 3	554-672	C-EU051379-IN	C-JF957684-TO C-JF957683-TO	Unknown	M**S**	$5.88x10^{-06}$
DNA-N 1	621-924	N-AY948438-IN	N-EU190970-IN N-AM418568-PK N-EF529519-PK N-EU391633-IN N-L41577-AU	Unknown	RGMCS**T**	$8.26x10^{-10}$
DNA-N 2	657-969	N-JF957690-TO N-JF957685-TO N-JF957686-TO	N-JF957693-TO N-JF957687-TO N-JF957688-TO	N-L41577-AU N-AM418568-PK N-EF529519-PK	RGMCS**T**	$2.56x10^{-07}$

		N-JF957691-TO	N-JF957689-TO	N-EU190970-IN		
		N-JF957692-TO	N-JF957695-TO	N-EU391633-IN		
		N-JF957694-TO	N-JF957696-TO			
DNA-N 3	805-844	N-JF957685-TO	N-FJ463047-CN	Unknown	**R**M	$7.00x10^{-04}$
		N-AM418568-PK	N-AF238878-CN			
		N-EF529519-PK	N-AF238879-CN			
		N-EU190970-IN	N-AY494787-CN			
		N-EU391633-IN	N-DQ826396-TW			
		N-JF957686-TO	N-EF470243-CN			
		N-JF957687-TO				
		N-JF957688-TO				
		N-JF957689-TO				
		N-JF957690-TO				
		N-JF957691-TO				
		N-JF957692-TO				
		N-JF957693-TO				
		N-JF957694-TO				
		N-JF957695-TO				
		N-JF957696-TO				
		N-L41577-AU				
DNA-N 4	400-439	N-FJ463047-CN	N-AF238879-CN	Unknown	RG	$4.92x10^{-03}$
		N-DQ826396-TW[T]	N-EF470243-CN			
DNA-N 5	225-264	N-DQ826396-TW	N-FJ463047-CN	N-L41577-AU	**G**	$2.95x10^{-02}$

2.6 References

Beetham, P. R., Hafner, G. J., Harding, R. M. & Dale, J. L. (1997). Two mRNAs are transcribed from banana bunchy top virus DNA-1. *J Gen Virol* 78, 229-236.

Bell, K. E., Dale, J. L., Ha, C. V., Vu, M. T. & Revill, P. A. (2002). Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Arch Virol* 147, 695-707.

Boni, M. F., Posada, D. & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176, 1035-1047.

Burns, T. M., Harding R. M. & Dale, J. L. (1995). The genome organisation of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76, 1471-1482.

Chen, L. I. F., Rojas, M., Kon, T., Gamby, K., Xoconostle Cazares, B. & Golbertson, R. L. (2009). A severe symptom phenotype in tomato in Mali is caused by a reassortant between a novel recombinant begomovirus (Tomato yellow leaf curl Mali virus) and a betasatellite. *Mol Plant Pathol* 10, 415-430.

Chen, Y. F., Chen, J. S., Zhang, H. R., Tang, X. S. & Du, Z. Y. (2007). Molecular evidence and sequence analysis of a natural reassortant between Cucumber mosaic virus subgroup IA and II strains. *Virus Genes* 35, 405-413.

Dale, J. L. (1987). Banana Bunchy Top - An economically important tropical plant-virus disease. *Adv Virus Res* 33, 301-325.

Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9, 267-276.

Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573-582.

Grigoras, I., Timchenko, T., Grande-Perez, A., Katul, L., Vetten, H. J. & Gronenborn, B. (2010). High variability and rapid evolution of a nanovirus. *J Virol* 84, 9105-9117.

Grigoras, I., Timchenko, T., Katul, L., Grande-Perez, A., Vetten, H. J. & Gronenborn, B. (2009). Reconstitution of authentic nanovirus from multiple cloned DNAs. *J Virol* 83, 10778-10787.

Gu, H., Zhang, C. & Ghabrial, S. A. (2007). Novel naturally occurring Bean pod mottle virus reassortants with mixed heterologous RNA1 genomes. *Phytopathology* 97, 79-86.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696-704.

Hafner, G. J., Harding, R. M. & Dale, J. L. (1997). A DNA primer associated with banana bunchy top virus. *J Gen Virol* 78, 479-486.

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S. & Robertson, D. (1999). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Crit Rev Plant Sci* 18, 71-106.

Harding, R. M., Burns, T. M., Hafner, G., Dietzgen, R. G. & Dale, J. L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *J Gen Virol* 74, 323-328.

Harkins, G. W., Martin, D. P., Duffy, S., Monjane, A. L., Shepherd, D. N., Windram, O. P., Owor, B. E., Donaldson, L., van Antwerpen, T., Sayed, R. A., Flett, B., Ramusi, M., Rybicki, E. P., Peterschmitt, M. & Varsani, A. (2009). Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. *J Gen Virol* 90, 3066-3074.

Heath, L., Van Der Walt, E., Varsani, A. & Martin, D. P. (2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* 80, 11827-11832.

Horser, C. L., Harding, R. M. & Dale, J. L. (2001a). Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *J Gen Virol* 82, 459-464.

Horser, C. L., Karan, M., Harding, R. M. & Dale, J. L. (2001b). Additional Rep-encoding DNAs associated with banana bunchy top virus. *Arch Virol* 146, 71-86.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. (2007). Reassortment and concerted evolution in Banana bunchy top virus genomes. *J Virol* 81, 1746-1761.

Hughes, A. L. (2004). Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. *Mol Phylogenet Evol* 30, 287-294.

Hyder, M. Z., Shah, S. H., Hameed, S. & Naqvi, S. M. S. (2011). Evidence of recombination in the Banana bunchy top virus genome. *Infect Genet Evol* 11, 1293-1300.

Jeske, H., Lütgemeier, M. & Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO Journal* 20, 6158-6167.

Jones, D. R. (1999). Introduction to Banana, Abaca and Enset. In *Diseases of banana, abaca and enset*. Edited by D. R. Jones.

Kagy, V., Thomas, J. E., Sharman, M. & Mademba-Sy, F. (2001). First record of banana bunchy top disease in New Caledonia. *Australas Plant Pathol* 30, 71-71.

Karan, M., Harding, R. M. & Dale, J. L. (1994). Evidence for two groups of banana bunchy top virus isolates. *J Gen Virol* 75, 3541-3546.

Karan, M., Harding, R. M., Dale, J. L. (1997). Association of Banana Bunchy Top Virus DNA Components 2 to 6 with Bunchy Top Disease. *Molecular Plant Pathology On-line*.

Kumar, P., Hanna, R., Alabi, O., Soko, M., Oben, T., Vangu, G. & Naidu, R. (2011). Banana bunchy top virus in sub-Saharan Africa: investigations on virus distribution and diversity. *Virus Res*.

Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009). Widely Conserved Recombination Patterns among Single-Stranded DNA Viruses. *J Virol* 83, 2697-2707.

Magee, C. J., Scientific, C. f. & Research, I. (1927). *Investigation on the bunchy top disease of the banana*: Council for Scientific and Industrial Research.

Maoka, T., Hayano, Y. S., Iwasaki, M., Yoshida, K. & Masuta, C. (2010). Mixed infection in tomato to ensure frequent generation of a natural reassortant between two subgroups of Cucumber mosaic virus. *Virus Genes* 40, 148-150.

Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562-563.

Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B., Vincent, C., Reynaud, B. & Lett, J. M. (2011). Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathog* 7.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462-2463.

Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrovir* 21, 98-102.

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., Martin, D. P. & Varsani, A. (2007). Successful application of FTA((R)) Classic Card technology and use of bacteriophage phi 29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *J Virol Methods* 140, 100-105.

Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218-225.

Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J. P., Jenny, C., Lebot, V., Risterucci, A. M., Tomekpe, K., Doutrelepont, H., Ball, T., Manwaring, J., de Maret, P. & Denham, T. (2011). Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc Natl Acad Sci U S A* 108, 11311-11318.

Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* 82, 655-665.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci U S A* 98, 13757-13762.

Qiu, W. P. & Moyer, J. W. (1999). Tomato spotted wilt tospovirus adapts to the TSWV N gene-derived resistance by genome reassortment. *Phytopathology* 89, 575-582.

Rohde, W., Randles, J. W., Langridge, P. & Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* 176, 648-651.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G. & Stanley, J. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* 293, 63-74.

Sharman, M., Thomas, J. E., Skabo, S. & Holton, T. A. (2008). Abaca bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch Virol* 153, 135-147.

Shepherd, D. N., Martin, D. P., Lefeuvre, P., Monjane, A. L., Owor, B. E., Rybicki, E. P. & Varsani, A. (2008). A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *J Virol Methods* 149, 97-102.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* 34, 126-129.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.

Thomas, J. E., Iskra-Caruana, M. L. & Jones, D. R. (2000). Diseases caused by viruses, bunchy top. *Diseases of banana, abacá and enset*, 241–283.

Timchenko, T., Katul, L., Aronson, M., Vega-Arreguin, J. C., Ramirez, B. C., Vetten, H. J. & Gronenborn, B. (2006a). Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. *J Gen Virol* 87, 1735-1743.

Timchenko, T., Katul, L., Aronson, M., Vega-Arreguín, J. C., Ramirez, B. C., Vetten, H. J. & Gronenborn, B. (2006b). Infectivity of nanovirus DNAs: Induction of disease by cloned genome components of Faba bean necrotic yellows virus. *J Gen Virol* 87, 1735-1743.

van der Walt, E., Martin, D. P., Varsani, A., Polston, J. E. & Rybicki, E. P. (2008). Experimental observations of rapid Maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virol J* 5.

Vetten, H. J., Chu, P. W., Dale, J. L., Harding, R. M., Hu, J., Katul, L., Kojima, M., Randles, J. W., Sano, Y. & Thomas, J. E. (2005). Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses, pp. 343-335. Edited by C. M. Fauquet, Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. San Diego, London: Academic Press.

Wanitchakorn, R., Harding, R. M. & Dale, J. L. (2000). Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch Virol* 145, 593-602.

Wardlaw, C. (1961). *Banana diseases, including plantains and abaca.*: Longmans, Green and Co. Ltd, London.

Xie, W. S. & Hu, J. S. (1995). Molecular cloning, sequence analysis, and detection of banana bunchy top virus in Hawaii. *Phytopathology* 85, 339-347.

Chapter 3

The global distribution of *Banana bunchy top virus* reveals little evidence for frequent recent, human-mediated long distance dispersal events

---

Contents

This work has been accepted by Virus Evolution and is presented in a similar manner to that of the manuscript:

Stainton, D., Martin, D., Muhire, B., Lolohea, S., Halafihi, M., Lepoint, P., Blomme, G., Crew, K. S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D. A., Mabvakure, B., Lemey, P., Harkins, G., Thomas, J. E., Varsani, A. The global distribution of *Banana bunch top virus* reveals little evidence for frequent recent, human-mediated long distance dispersal events (In Print)

3.1. Abstract

Banana bunchy top virus (BBTV; family *Nanoviridae*, genus *Babuvirus*) is a multi-component single-stranded DNA (ssDNA) virus which infects banana plants in many regions of the world, often resulting in large scale crop losses. We analysed 171 banana leaf samples from 14 countries and recovered, cloned and sequenced 855 complete BBTV components including 94 full genomes. Importantly, full genomes were determined from eight countries where previously no full genomes were available (Samoa, Burundi, Republic of Congo, Democratic Republic of Congo, Egypt, Indonesia, the Philippines and the United States [Hawaii]). Accounting for recombination and genome component reassortment, we examined the geographic structuring of global BBTV populations to reveal both that the global hotspots of BBTV diversity are Southeast Asia / Far East and India, and that BBTV populations circulating elsewhere in the world have all potentially originated from single founder events. Most importantly, we find no convincing evidence of frequent modern inter-continental movements of BBTV variants to different parts of the world.

3.2. Introduction

Bananas are grown in over 130 countries and are ranked fourth, after wheat, rice and maize, in importance as a food crop in the world (FAOSTAT; Perrier *et al.*, 2011). Domesticated bananas are thought to have originated somewhere in the vicinity of New Guinea, Indonesia, the Philippines, or the Southeast Asia Peninsula (Perrier *et al.*, 2011) between 7,000 and 10,000 years ago (Denham *et al.*, 2003). Banana cultivation subsequently spread to other parts of the world reaching Cameroon in West Africa and the Indian Ocean island of Madagascar possibly as early as 3,000 years ago. During the period between 1,500 and 700 years ago, different banana varieties were likely introduced and reintroduced to Africa and the south-west Indian Ocean Islands many times (Lejju *et al.*, 2006; Randrianja & Ellis, 2009).

Banana bunchy top disease (BBTD) is one of the most important diseases of banana, causing severe crop losses in many banana-growing regions outside the Americas. Banana plants apparently displaying BBTD symptoms were described in Fiji as early as 1880s (Magee, 1927). In the 1930s the banana aphid, *Pentalonia nigronervosa*, was found to transmit the disease in a persistent manner (Magee, 1940). However, it was not until the 1990s that an icosahedral single-stranded DNA virus with six genome components was identified as the causative agent. This virus, BBTV (Burns *et al.*, 1995; Burns *et al.*, 1994; Harding *et al.*, 1991; Harding *et al.*, 1993; Thomas & Dietzgen, 1991), is now recognised as the type member of the genus *Babuvirus* in the family *Nanoviridae*.

The six genome components of BBTV are each approximately 1,000 nucleotides (nt) long and are named DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C, and DNA-N (formerly DNA-1 to DNA-6 respectively) (King *et al.*, 2012). DNA-R encodes a replication-associated protein (Rep), DNA-S a capsid protein (CP), DNA-M a movement protein (MP), DNA-C a cell-cycle link protein (Clink) and DNA-N a nuclear shuttle protein (NSP) (Aronson *et al.*, 2000a; Hafner *et al.*, 1997b; Wanitchakorn *et al.*, 2000a; Wanitchakorn *et al.*, 2000c). The function of DNA-U3 is currently unknown. All components of individual viruses contain two sequence elements which are highly similar across the components: a common region stem-loop (CR-SL) element and a common region major (CR-M) element (Burns *et al.*, 1995). The CR-SL is involved in replication and contains both a hairpin structure with a highly conserved nonanucleotide sequence (TATTATTAC), and three repeated 5 nt long sequences, called iterons, that are likely involved in the recognition and/or binding of *rep* to the virion

strand origin of replication (*v-ori*) (Burns *et al.*, 1995; Herrera-Valencia *et al.*, 2006). The CR-M is thought to be involved in transcription (Burns *et al.*, 1995) and also contains most of the binding sites for a primer molecule that is involved in complementary strand DNA synthesis (Hafner *et al.*, 1997a).

Components of BBTV isolates broadly fall into two geographically well-defined phylogenetic groups, the South Pacific group (SPG) and the Asian group (AG) (Karan *et al.*, 1994). Despite these phylogenetic groups having been defined based on the geographic origins of genomic component sequences available in the mid-1990s, subsequently determined BBTV sequences have continued to phylogenetically cluster within one or the other of these groups, with almost all sequences sampled outside of Southeast Asia falling into the SPG. Although the SPG and AG have also been respectively referred to as the Pacific/Indian Ocean (PIO) and the Southeast Asia (SEA) groups (Yu *et al.*, 2012), here we will continue to use their original names.

It is likely that this geographic structuring has arisen because the rates of natural and/or human-mediated long-distance BBTV movement have been low enough for geographically separated populations of these viruses to have differentiated from one another. It remains unknown, however, whether the current geographical distribution of BBTV variants arose concomitantly with the slow, pre-historic spread of banana cultivation across the Pacific, the Indian Ocean, Asia and Africa, or whether it is a consequence of poorly-regulated agricultural trade during the modern globalisation era. It is additionally plausible that the current distribution of BBTV might be due to a combination of both these factors. Importantly, the degrees of geographic structure evident within contemporary genomic sequence data might be high enough to yield insights into when, and from where the BBTV populations in particular continents, countries or districts were founded. Such insights would be extremely valuable in determining, for example, whether modern movements of banana germplasm across the globe have had an appreciable impact on BBTV distributions.

The potential for human-mediated dissemination of BBTV is high since cultivated bananas are sterile and are propagated vegetatively. Also, a banana plant infected with BBTV can take between 25 and 85 days to develop visible symptoms (Hooks *et al.*, 2008) meaning that infected but symptomless banana propagules could be inadvertently transferred to regions where *P. nigronervosa* is present, and the BBTV variants within these propagules might be successfully transmitted and establish new BBTV populations.

It is also likely that, as is the case with other related ssDNA viruses (Duffy & Holmes, 2008; Duffy *et al.*, 2008; Firth *et al.*, 2009; Grigoras *et al.*, 2010; Harkins *et al.*, 2009; Harkins *et al.*, 2014; Kraberger *et al.*, 2013; van der Walt *et al.*, 2008), BBTV is evolving at a sufficient rate that evidence of such movement events should be encoded within the phylogenetic relationships of genomic component sequences sampled from extant BBTV populations.

Phylogenetic inference of BBTV movement dynamics might, however, be confounded by two other evolutionary processes that occur in BBTV, genome component reassortment and homologous recombination. Due to each genome component being packaged individually into separate virions, new infections that are propagated from mixed BBTV infections will frequently contain an assortment of different genome components. BBTV isolates that have genome components derived from two or more different parental viruses have been inferred using a variety of phylogenetic (Hu *et al.*, 2007; Yu *et al.*, 2012), and statistical recombination detection methods (Martin *et al.*, 2010; Stainton *et al.*, 2012). Similar examples of component reassortment have also been found in a number of other nanovirus species (Grigoras *et al.*, 2014; Hu *et al.*, 2007; Pita *et al.*, 2001; Savory & Ramakrishnan, 2014).

The known sequences of many of individual BBTV genome components also carry evidence of homologous recombination (Banerjee *et al.*, 2014; Hyder *et al.*, 2011; Stainton *et al.*, 2012; Wang *et al.*, 2013). While the accuracy of phylogenetic reconstructions for individual genome components could be significantly undermined by homologous recombination, both recombination and reassortment will undermine the accuracy of full-genome phylogenetic reconstructions (Posada & Crandall, 2002; Schierup & Hein, 2000).

In order to both gain a more detailed view of global BBTV diversity and assess the geographical structuring of BBTV populations at higher resolution than has previously been achievable, we determined the sequences of 855 full BBTV genome components from samples collected from across much of the known BBTV geographic range (Figure 3.1, Supplementary Table 3.1). Accounting for recombination and reassortment we find that the diversity and phylogeographic structure of contemporary known BBTV populations is entirely consistent with these populations having been founded in different parts of the world by single, and possibly ancient, BBTV movement events.

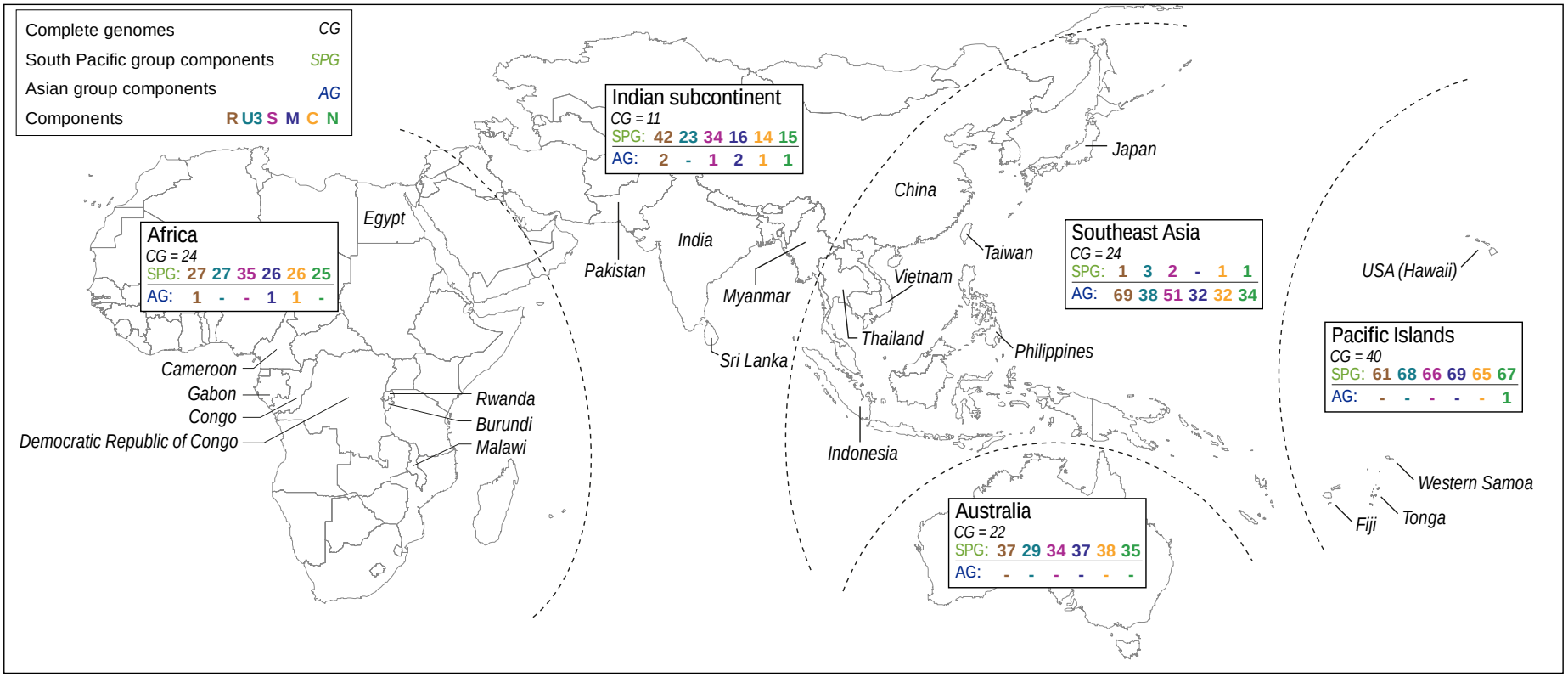


Figure 3.1: Geographical distribution of BBTV isolates. Summaries of component numbers and full genomes are provided for different regions. Accession numbers and specific component information can be found in Figure 3.2 and Supplementary Table 3.1.

3.3. Materials and methods

3.3.1 Extraction and sequencing

Samples were collected from 171 banana plants displaying stunting, bunched leaves, and Morse-code like streaking between leaf margins and the midrib: all of which are symptoms characteristic of BBTD. Samples were collected between 1989 and 2012 from Australia (n=40 isolates), four African countries (n=23 isolates), three Pacific island groups (n=69 isolates), two Indian Subcontinent countries (n=8 isolates), and four Southeast Asia / Far East countries (n=31 isolates), summarised in Figure 3.1 and Supplementary Table 3.1. Banana material in this study was collected by; TOS2-93, by DS, SL, MH, SK, MW, DAC, AV and students from Tonga College, BU1-20 by GB and PL, B2817-47 by KSC, MS, JET. All other banana material used in this study was sourced from a worldwide collection stored at the University of Queensland, managed by JET.

DNA extractions, amplification and sequencing of BBTV genome components were carried out as described previously (Stainton *et al.*, 2012). Briefly, sampled leaf material (fresh or dried) was homogenised and total DNA was extracted using an GenCatch Plant Genomic DNA Purification kit (Epoch Biolabs, USA). Circular DNA was preferentially enriched using the TempliPhi amplification kit (GE Healthcare, USA) as described previously (Owor *et al.*, 2007; Shepherd *et al.*, 2008). BBTV genome components were PCR amplified using component-specific back-to-back primers described in Stainton *et al.* (2012). The resulting amplicons were resolved on an agarose gel, gel purified, cloned and single transformed plasmid clones were sequenced at Macrogen Inc (South Korea). Sequence contigs were assembled using DNA Baser Sequence Assembler v4 (Heracle Biosoft SRL, Romania). Where possible all six components were sequenced from each sample, although we were unable to recover all of the components from some samples (Figure 3.2, Supplementary Table 3.1).

3.3.2 Datasets

All full components sequenced as part of this study, along with all full BBTV component sequences available in GenBank (downloaded 1st March 2014, see Supplementary Table 3.1 for isolate information) were split into individual component specific datasets (CSD) all starting at the ‘TATTAC’ region of the nonanucleotide sequence motif. These sequences were aligned using MUSCLE (Edgar, 2004) implemented in MEGA5 (Edgar, 2004; Tamura *et al.*, 2011). Aligned component sequences identified from the same sample were then

concatenated into a single sequence. As outlined in Stainton *et al.* (2012), blank sequences (i.e. composed entirely of "-"characters) were used where component sequences were not available both to maintain the component order and for alignment purposes. These concatenated sequences were labelled as the concatenated dataset (CD) (Supplementary Table 3.1 contains information on cognate BBTV components). From this dataset, a new BBTV genome dataset called the full genome dataset (FGD) was created containing all available full genome sequences (all six components sequenced). Except where stated, *Abaca bunch top virus* (ABTV) sequences were used as an out group for all datasets. Following recombination and reassortment analysis, a recombination-free CSD (RF-CSD) for each component and a recombination and reassortment-free FGD (RF-FGD) were constructed from the CSDs and FGD respectively (see below for recombination and reassortment details).

3.3.3 Pairwise nucleotide sequence identity analyses

Percentage pairwise nucleotide identities of complete BBTV genomes (in the context of the FGD) and of individual components (in the context of the CSDs) were determined using Sequence Demarcation Tool (SDT) v1.2 (Muhire *et al.*, 2014) with the MUSCLE-based alignment option. CSDs and the FGD were all analysed with SDT without accounting for recombination. Distributions of pairwise nucleotide identities of the FGD and CSDs were used to tentatively classify BBTV genomes into groups and subgroups based on the majority of the components in a similar way to Varsani *et al.* (2014).

All CSD were split into AG and SPG based on Neighbor-joining trees reconstructed using the Jukes-Cantor model as implemented in MEGA 5 (Tamura *et al.*, 2011) (data not shown), and percentage pairwise identity was calculated for each group using SDT v1.2.

Percentage pairwise identities were determined for the FGD and CSD sequences for five geographic regions: Africa (Burundi, Cameroon, Congo, Democratic Republic of Congo, Egypt, Gabon, Malawi, Rwanda), the Indian Subcontinent (India, Myanmar, Pakistan, Sri Lanka), Southeast Asia / Far East Asia (China, Indonesia, Japan, Philippines, Taiwan, Thailand, Vietnam), Pacific Islands (Fiji, Hawaii, Kingdom of Tonga, Samoa), and Australia.

3.3.4 Recombination and reassortment analyses

All recombination and reassortment events were detected using RDP4.27 (Martin *et al.*, 2010), a recombination detection programme which implements the following detection methods: RDP (Martin & Rybicki, 2000), GENECONV (Padidam *et al.*, 1999), Bootscan

(Martin *et al.*, 2005), Maxchi (Smith, 1992) Chimera (Posada & Crandall, 2001), SiScan (Gibbs *et al.*, 2000), and 3Seq (Boni *et al.*, 2007). Recombination events were considered credible when, along with phylogenetic evidence, an event was identified by at least three detection methods with an associated p-value < 0.05 and with at least one method having an associated p-value < 0.001. Reassortment events were considered credible when, along with phylogenetic evidence, an event was identified by at least two detection methods with an associated p-value < 0.05, with at least one method having an associated p-value < 0.001. Intra-component recombination events were identified using the single component specific datasets (CSD). Reassortment events were identified using the CD. Specifically, recombination events identified by RDP4.27 that had associated breakpoints which spanned an entire component were identified as reassortment events.

Due to the large number of sequences being analysed, as well as issues with accurately aligning all six components, a dataset containing all sequences from all components was not used to detect evidence of possible inter-component recombination. Therefore, all intra-component recombinant regions with an unknown minor parent were further analysed using BLASTn (Altschul *et al.*, 1990) to determine whether transferred sequence fragments identified as having unknown origins could have credibly been derived from different BBTV components.

3.3.5 Phylogenetic analysis of BBTV geographic distributions

Maximum-likelihood (ML) phylogenetic trees were constructed using the recombination-free CSDs with PhyML 3 (Guindon *et al.*, 2010) applying the best fit nucleotide substitution model for each dataset determined using jModelTest (Posada, 2008) with 100 bootstrap replicates to determine branch support. For the recombination and reassortment-free FGD a ML phylogenetic tree was constructed using RAxML (Stamatakis, 2014) with 100 bootstrap replicates. All phylogenetic trees were rooted with ABTV sequences and branches with <60% bootstrap support were collapsed using Mesquite v2.75 (http://mesquiteproject.org/). RAxML was used for these particular trees rather than PHYML because it has been specifically optimised to construct phylogenetic trees from sequences containing large amounts of missing data (Izquierdo-Carrasco *et al.*, 2011).

3.3.6 Selection analysis

The open reading frames (ORFs) of DNA-R, DNA-S, DNA-M, DNA-C, and DNA-N were identified for all BBTV sequences in this study, along with those of *Faba bean necrotic yellows virus* (FBNYV) (downloaded 10th March 2014) and CBDV (downloaded 10th July 2014) sequences with full ORFs. ABTV sequences were not analysed due to the small number (n=13) of component sequences available in public databases. The ORFs were codon aligned using MUSCLE (Edgar, 2004) implemented in MEGA5 (Tamura *et al.*, 2011). The *rep* genes (on DNA-R) of BBTV, FBNYV and CBDV isolates were aligned together. Due to a lack of similarity of all FBNYV genes other than *rep* with those of CBDV and BBTV, all other FBNYV ORFs were aligned and analysed independently. The DNA-S, DNA-M, DNA-C and DNA-N genes of BBTV and CBDV were aligned together. For each species all aligned datasets were analysed separately for selection using the HyPhy package (Pond *et al.*, 2005), with MEME (Murrell *et al.*, 2012) being used to identify episodic and pervasive positive selection and FUBAR (Murrell *et al.*, 2013) being used to identify pervasive positive and negative selection. Sites which showed selection were then compared to identify whether selection signals at particular sites were conserved between species.

3.4. Results and discussion

3.4.1 Sample collection and sequencing

Although BBTV populations seriously constrain banana production throughout much of the eastern hemisphere, the worldwide genetic diversity of BBTV remains poorly understood. We therefore amplified, cloned and sequenced 855 complete BBTV components (DNA-R, n=137; DNA-U3, n=138; DNA-S, n=146; DNA-M, n=146; DNA-C, n=143; DNA-N, n=145) from 171 BBTV infected banana plants from 14 countries spanning the known geographical range of this virus (Figure 3.2, Supplementary Table 3.1, accession numbers KM607005–KM607859).

A subset of the newly determined genome component sequences constitute 94 complete BBTV genomes (i.e. instances where all six components have been sequenced from a single sample). These 94 genomes include those sampled in countries/territories from which either no BBTV sequence data was previously available (Congo and Samoa) or for which no full genomes have previously been sequenced (Burundi, Democratic Republic of Congo, Egypt, Indonesia, the Philippines and Hawaii). This new sequence data more than doubles the number of publically available BBTV full genome component sequences. All GenBank

accession numbers for these and other publically available BBTV sequences used in this study can be found in Supplementary Table 3.1.

In total 1,191 BBTV and 13 ABTV component sequences (ABTV DNA-M n=3, all other components n=2) were assembled into a CD and CSD DNA-R (n=242), DNA-U3 (n=190), DNA-S (n=225), DNA-M (n=186), DNA-C (n=180), DNA-N (n=181). These sequences have collectively been recovered from a total of 318 plant samples (171 in this study) from 25 countries (14 sampled in this study; Figure 3.2 and Supplementary Table 3.1). A FGD containing isolates with all six component sequences was assembled from the CD and contained 121 full BBTV genomes and two full ABTV genomes.

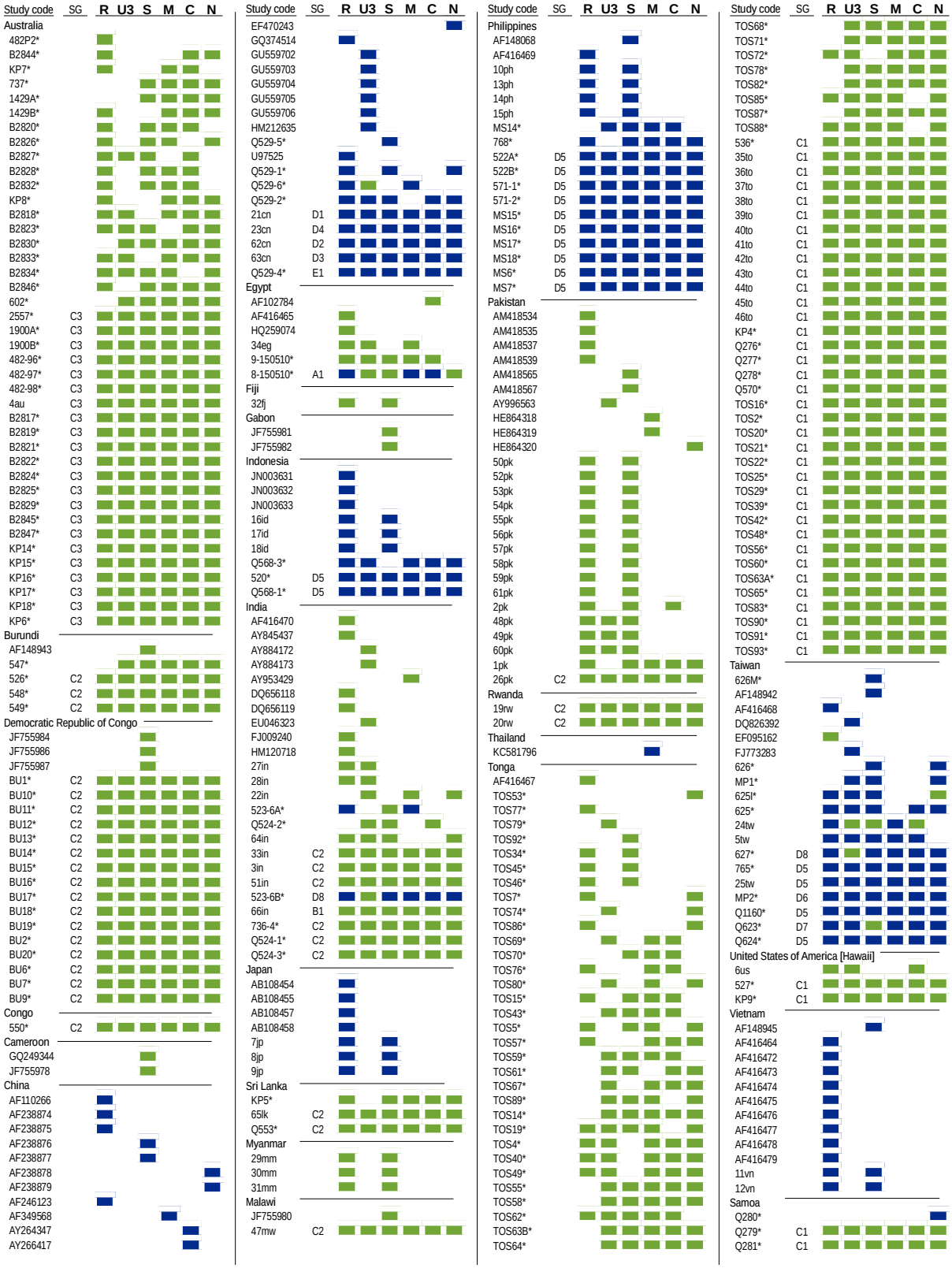


Figure 3.2: Overview of sequenced BBTV components. Isolates are grouped by country, with each individual component depicted with a blue or green square depending on whether the component falls phylogenetically into the AG (blue) or SPG (green) group (phylogenetic tree not shown). Isolates with asterisks were sequenced as part of this study. Full genome subgroups (SG), based on percentage pairwise identities, are shown for all isolates with all six components sequenced. Full isolate information, including accession numbers and references, are in Supplementary Table 3.1.

3.4.2 Classification of the genome segments and full genomes

The DNA-U3 components were most diverse, sharing >74% pairwise identity followed by DNA-S and DNA-M both with >82%, then DNA-N, DNA-C and DNA-R with >83%, >85% and >88% pairwise identities, respectively. Collectively the segments in the FGD shared >85% pairwise identity. For the FGD, the genome sequences which shared >85% but <94% pairwise identity were subdivided into groups A to E. Within these groups, genomes with >98% pairwise identity were further divided into subgroups A1, B1, C1-3, D1-8 and E1 (Figure 3.2; Supplementary Table 3.1).

With the exception of DNA-S, the genetic diversity amongst the currently sampled AG genome components is generally greater than that amongst the corresponding SPG components: DNA-R (AG>91%, SPG>94%), DNA-U3 (AG>76%, SPG>81%), DNA-M (AG>89%, SPG>91%), DNA-C (AG>89%, SPG>94%), and DNA-N (AG>89%, SPG>91%) components. In the case of DNA-S, the SPG sequences are >87% identical whereas the AG sequences are >92% identical.

The percentage pairwise identities of genome components sampled from five major regions of the world (Africa, the Indian subcontinent, Southeast Asia / Far East, the Pacific Islands and Australia) indicated that the greatest degree of BBTV sequence diversity occurs within the Southeast Asia / Far East /Indian subcontinent regions (Table 3.1). This was true for the full genome dataset and the majority of the individual components (DNA-R, DNA-U3, DNA-S, DNA-C). The significant diversity observed in Africa is contributed mainly by the AG-like DNA-R, -M and -C components of isolate 8-150510 (Figure 3.2). This suggests that the true global diversity of BBTV could be best inferred by increased sampling effort in these regions. Unlike the other components, DNA-M diversity is greatest within the Africa region, and DNA-N diversity is highest within the Pacific Island region.

Table 3.1: Percentage pairwise identities of individual BBTV genome components that have been sampled from different geographical regions.

Full genome / component	Region	Percentage pairwise identity
Full genome	Africa	>91.3
	Australia	>97.3
	Indian Subcontinent	>87.1
	Pacific islands	>97.3
	Southeast Asia	>90.1
DNA-R	Africa	>90.1
	Australia	>98.6
	Indian Subcontinent	>89.5
	Pacific islands	>95.5
	Southeast Asia	>89.1
DNA-U3	Africa	>87.2
	Australia	>96.7
	Indian Subcontinent	>82.5
	Pacific islands	>91.3
	Southeast Asia	>75.9
DNA-S	Africa	>96.8
	Australia	>88.9
	Indian Subcontinent	>87.8
	Pacific islands	>96.9
	Southeast Asia	>88.2
DNA-M	Africa	>82.1
	Australia	>92.4
	Indian Subcontinent	>83.4
	Pacific islands	>91.8
	Southeast Asia	>89.2
DNA-C	Africa	>86.7
	Australia	>97.6
	Indian Subcontinent	>86.4
	Pacific islands	>97.5
	Southeast Asia	>86.3
DNA-N	Africa	>97.2
	Australia	>98.9
	Indian Subcontinent	>85.0
	Pacific islands	>83.8
	Southeast Asia	>84.7

3.4.3 Reassortment analyses

Given that BBTV genome components are individually encapsidated, mixed infections will often result in genome component reassortment (Hu *et al.*, 2007; Stainton *et al.*, 2012). To ensure the accuracy of our FGD phylogenetic analyses it was vital that we identified and removed from our datasets genome components that had been acquired by reassortment. Towards this end the CD was analysed for evidence of reassortment using RDPv4.27 (Martin *et al.*, 2010), with manual identification of reassortment events as detected recombination events that had inferred breakpoint locations spanning entire components (Stainton *et al.*, 2012). Given that this analysis involved almost four times more full genomes than previous BBTV reassortment analyses, it is not surprising that of the 75 isolates detected as reassortants, only ten had been detected previously (Hu *et al.*, 2007; Stainton *et al.*, 2012).

These 75 isolates carried evidence of 40 different reassortment events (Figure 3.3, Supplementary Table 3.2). All components were represented amongst these events, albeit with some components having been transferred more than others. Component DNA-U3 was found to be the most commonly transferred component (11 events), followed by DNA-M (8 events), DNA-S and DNA-N (both with 7 events), DNA-C (5 events) and DNA-R (2 events).

Similar reassortment analyses in *Cardamom bushy dwarf virus* (CBDV) and viruses in the genus Nanovirus which lack a DNA-U3 component have also found that DNA-M and/or DNA-N are among the most frequently transferred nanovirus components during reassortment (Grigoras *et al.*, 2014; Savory & Ramakrishnan, 2014). However, DNA-U3, which is only present in Babuviruses (BBTV, ABTV and CBDV), was not found to be amongst the most frequently transferred genome components during CBDV reassortment (Savory & Ramakrishnan, 2014). This suggests that patterns of component transfer are not absolutely conserved between different species.

Of the 75 reassortant genomes that we detected, 34 had one detectable reassortment event, 33 had two and eight had three. Overall, ~38% of all isolates with at least three sequenced components (75/196) show evidence of at least one component having been acquired by reassortment. Crucially, 12 of the 40 reassortment events were each detected in multiple genomes. This strongly suggests that these events occurred in an ancestor of these genomes

and therefore that reassortment yielded viable viruses that went on to become epidemiologically relevant.

Our detection of reassortment events between AG and SPG genomes sampled in Egypt, China, India and Taiwan (Figure 3.2) is consistent with the geographic range of the AG and SPG lineages overlapping in these regions. This overlap suggests that the Indian / Southeast Asian / Far Eastern region is likely the geographic hotspot of BBTV diversity and might even be the region where the most recent common ancestor of all currently sampled BBTV isolates originated.

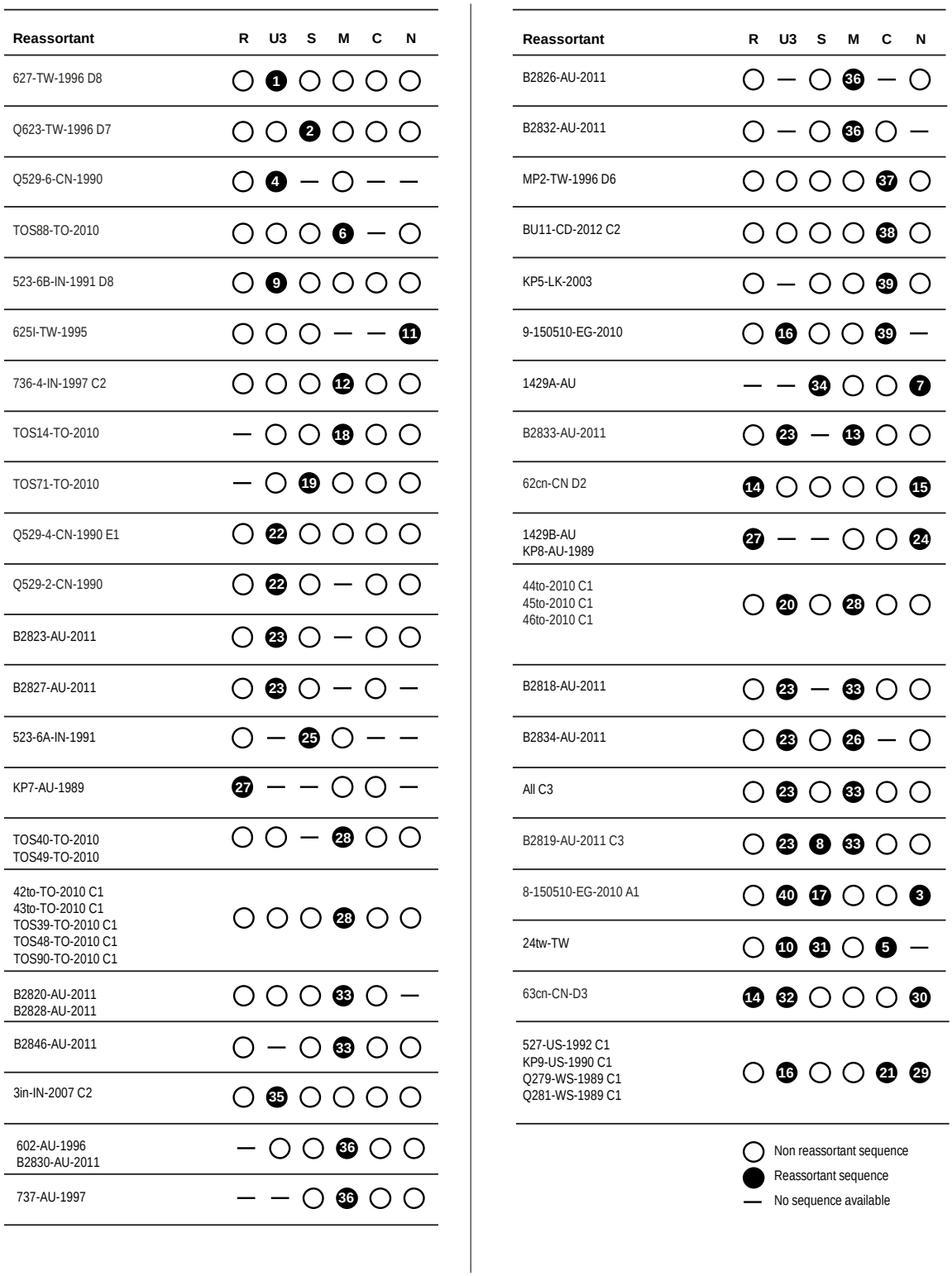


Figure 3.3: Detected reassortment events. As not all reassortant isolates consist of full genomes, circles depict component sequences which are available and a dash indicates where no component sequence is available. Components are shown as either non-reassortant sequences (white filled circles) or as reassortant sequences (black circles) with the corresponding reassortant event number. Further information on reassortment events can be found in Supplementary Table 3.2.

3.4.4 Recombination analyses

A number of studies have identified potential recombination events in BBTV (Banerjee *et al.*, 2014; Fu *et al.*, 2009; Hyder *et al.*, 2011; Islam *et al.*, 2010b; Stainton *et al.*, 2012), and, as with reassortment, it was important to account for these events during our subsequent phylogenetic analyses. We analysed the CSDs to identify recombinant sequences, the locations of recombination breakpoints, and the identities of likely parental viruses.

These analyses revealed that all components displayed at least some evidence of recombination (Figures 3.4 and 3.5; Supplementary tables 3.3-3.8), with the greatest number of recombination events being detected in DNA-U3 (12 events) and the fewest in DNA-M (two events). All components carried evidence of recombinant regions involving the CR-SL region (with breakpoints falling within and/or on either side of this region) but only DNA-U3, and -N have recombination regions involving the CR-M, all of which had breakpoints falling on either side of this region. Of the 18 recombination events that were identified within multiple isolates, nine are seen within isolates from multiple countries. As with the reassortment events that are observed in multiple different genomes, these recombination events apparently occurred within genomes that were ancestral to two or more of the sequences analysed here and indicate that at least some BBTV recombinants are epidemiologically relevant.

Twenty-two recombination events were detected within the components encoding genes of known function, DNA-R, -M, -N, -S and -C. As has been found in previous nanovirus recombination studies (Grigoras *et al.*, 2014; Hyder *et al.*, 2011; Savory & Ramakrishnan, 2014; Stainton *et al.*, 2012), we detected a similar numbers of recombination breakpoints within the non-coding and coding regions (24 and 20 breakpoints, respectively). In total 13 events resulted in recombinant genes that could express chimeric proteins. However, all 13 of these events involved recombination between closely related BBTV variants, meaning that these recombination events would have had only a minimal impact on encoded protein amino acid sequences (Lefeuvre *et al.*, 2009). Another possible sign of protein coding sequences having an impact on recombination patterns in BBTV is that the DNA-U3 component, which has no confirmed protein coding function, has a higher concentration of detectable recombination breakpoints than those of the known protein coding genes of other components. Interestingly DNA-U3 is also the component that appears to be most frequently exchanged by reassortment in BBTV. High frequencies of recombination in this component

might reflect the fact that it is mostly evolving neutrally with no risk that recombinants might express defective chimeric proteins (Lefeuvre *et al.*, 2009) and that there is therefore little conservation of epistatic interactions within this component.

Eighteen of the detected recombination events apparently involved the acquisition by BBTV isolates of genetic material derived through either inter-component recombination, or through recombination with non-BBTV babuvirus species (DNA-R, n=2; -U3, n=8; -S, n=3; -C, n=2; -N, n=3) (see Supplementary tables 3.3-3.8 for details). All were analysed using BLASTn (Altschul *et al.*, 1990), with four of the recombination events - U7 (in DNA-U3), S1 (in DNA-S), C2 (in DNA-C) and N3 (in DNA-N) - having potentially involved inter-component sequence transfers. BLASTn (Altschul *et al.*, 1990) analysis of the U7 recombinant region indicated that this had likely involved a BBTV satellite (accession # EU366175): a finding that corroborates the results in Fu *et al.* (2009). Whereas BLASTn analyses of events S1 and C2 indicated that the most likely sources of the recombinationally acquired sequences were BBTV DNA-M components, and event N3 (which was detected in ABTV) likely involved a sequence transfer from an ABTV DNA-S component.

Our analyses indicated the remaining 14 detected recombination events with unknown parents likely involved homologous recombination between BBTV and viruses belonging either to currently unsampled babuvirus species, or to divergent currently unsampled BBTV strains. This suggests that there may exist a far greater diversity of BBTV-like babuvirus species (or perhaps divergent BBTV strains) than is presently known. Also, the fact that recombination events that are inferred to involve currently unsampled babuvirus species are primarily evident in BBTV isolates sampled in Southeast Asia / Far East region (9/14 events) further suggests that this region is likely a major hotspot of ongoing recombination-driven BBTV diversification.

For recombination to occur between any particular pair of viruses the viruses must have overlapping geographic ranges, host ranges and cell tropisms. A number of plants, which are also hosts of *Pentalonia* spp. (*P. caladii* and *P. nigronervosa*), have been suggested as potential alternative hosts for BBTV including *Canna indica* (canna lily), *Hedychium coronarium* (white ginger lily) and *Colocasia esculenta* (taro) (Duay *et al.*, 2014; Foottit *et al.*, 2010). BBTV can be transmitted by *Pentalonia* spp. from an infected banana into *C. esculenta* (asymptomatic) and then back into a healthy banana plant to cause disease (Ram & Summanwar, 1984). *C. indica* and *H. coronarium* have also shown weak to moderate

reactions in BBTV-specific ELISA tests (Su *et al.*, 1992). Although Pinili *et al.* (2013) reported the successful transmission of an Okinawan BBTV isolate to *C. indica, C. esculenta* and *Alpinia zerumbet*, further studies have failed to confirm that these species are suitable hosts for other BBTV strains (Geering & Thomas, 1997; Hu *et al.*, 1996; Manickam *et al.*, 2002).

Regardless of the actual BBTV host-range, our results indicate that an increased sampling effort targeting uncultivated species in Southeast Asia / Far East and possibly India, may lead to the identification of both alternative BBTV host species and numerous other epidemiologically relevant babuvirus species. Indeed, the only other known babuvirus species have been identified from this region; CBDV from India (Mandal *et al.*, 2013) and ABTV from the Philippines and Malaysia (Sarawak) (Sharman *et al.*, 2008a).

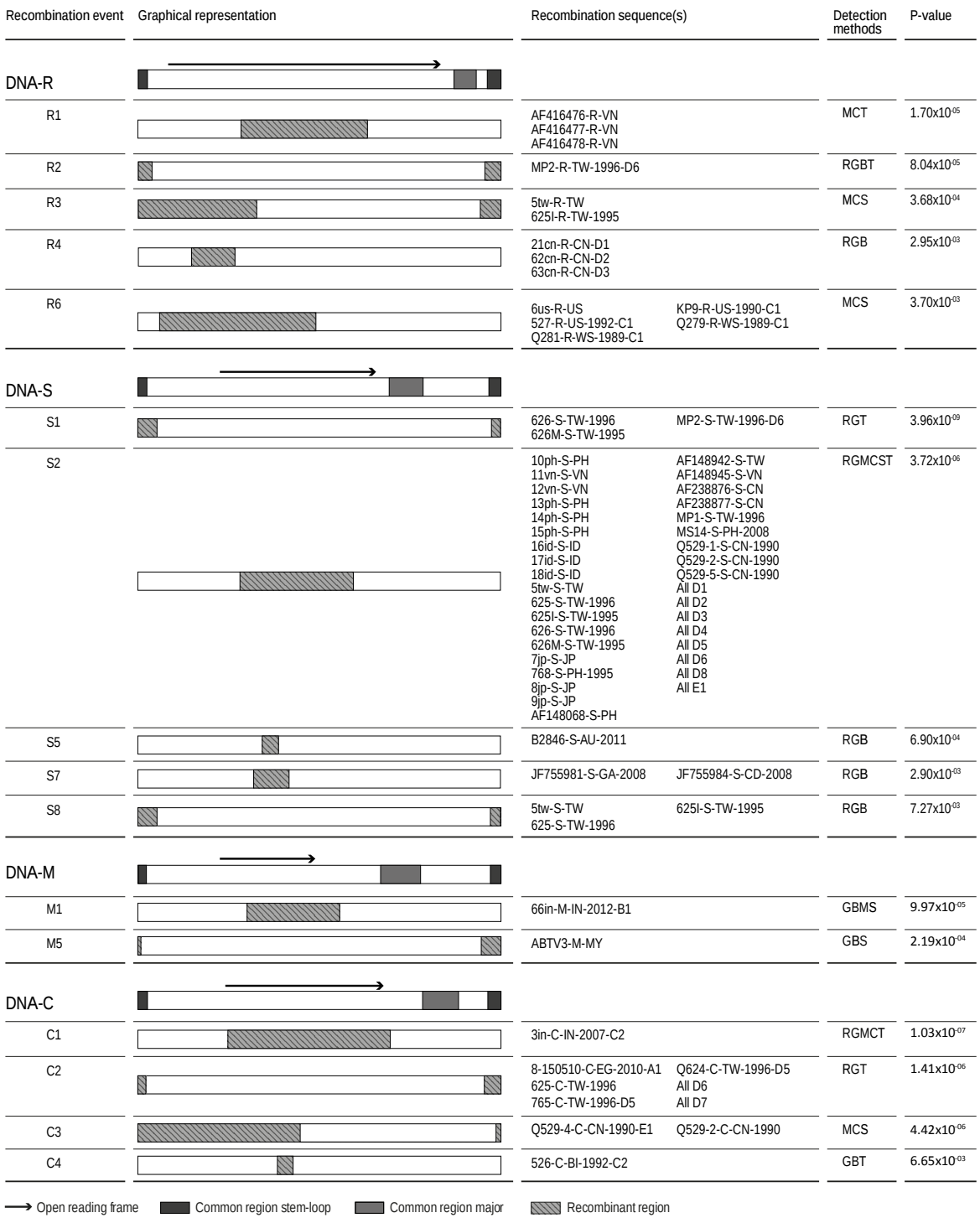


Figure 3.4: Recombination events detected in DNA-R, DNA-S, DNA-M and DNA-C. Methods which detected the event are shown by abbreviations: R -RDP; G -GENCONV; B -BOOTSCAN; M -MAXCHI; C -CHIMERA; S -SISCAN; T -3SEQ. The most significant P-value is shown with the detection method marked in bold. Further information on recombination events can be found in Supplementary tables 3.3, 3.5, 3.6, 3.7.

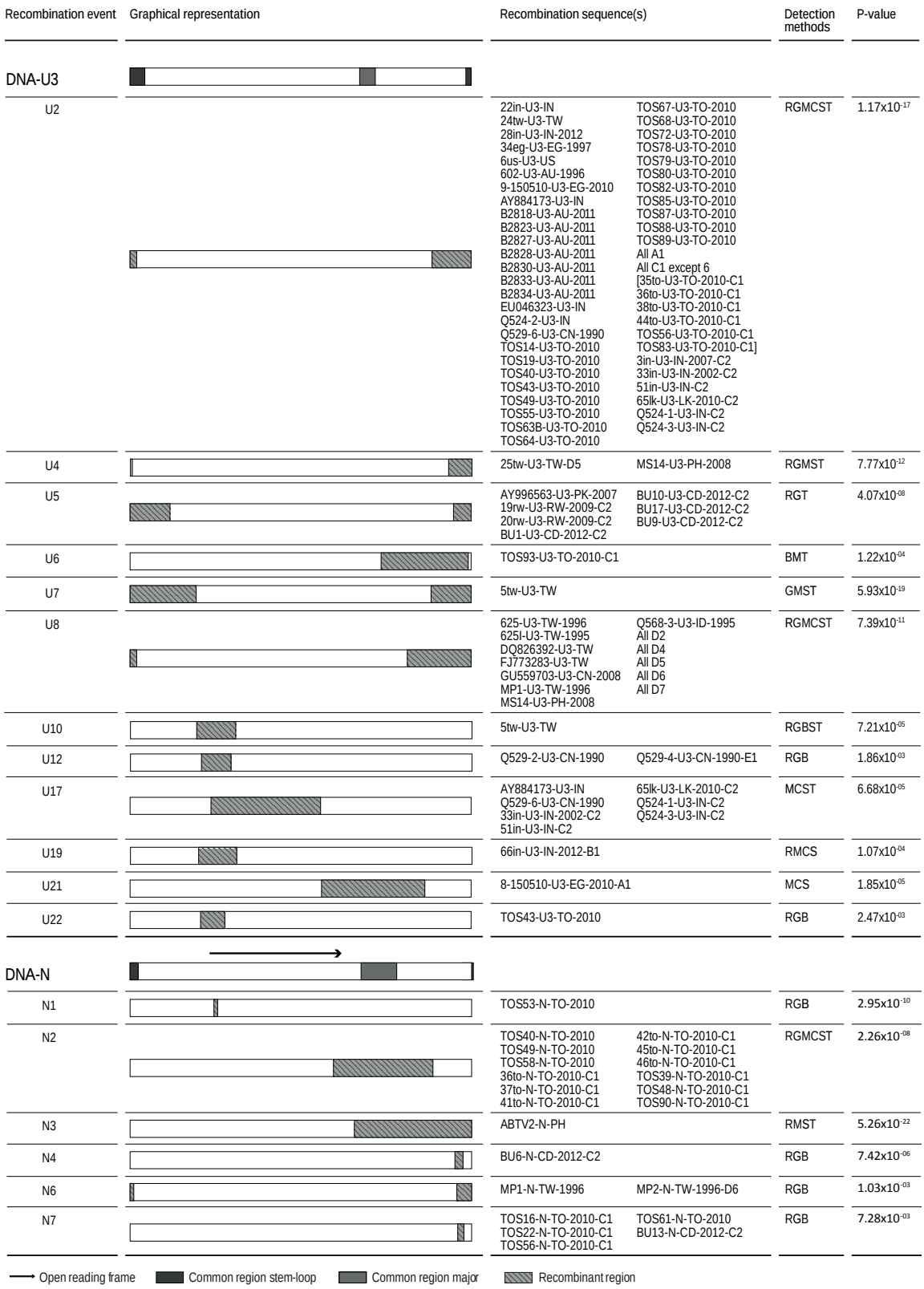


Figure 3.5: Recombination events detected in DNA-U3 and DNA-N. Methods which detected the event are shown by abbreviations: R -RDP; G -GENCONV; B -BOOTSCAN; M -MAXCHI; C -CHIMERA; S -SISCAN; T -3SEQ. The most significant P-value is shown with the detection method marked in bold. Further information on recombination events can be found in Supplementary tables 3.4 and 3.8.

3.4.5 Analysis of geographical structure within BBTV phylogenies

Banana domestication is thought to have occurred on the Southeast Asian peninsula or its adjacent islands between seven and ten thousand years ago (Denham *et al.*, 2003; Perrier *et al.*, 2011). Given the potential for human-mediated dissemination of BBTV via the subsequent worldwide movements of infected banana propagules, we aimed to examine the degree to which the phylogenetic relationships between BBTV isolates reflected their geographic origins. After removing genome components in the FGD dataset that had been derived through reassortment and fragments of components that had been derived through recombination in both the FGD and CSDs (to respectively yield recombination free datasets; RF-FGD and RF-CSDs), we constructed ML phylogenetic trees for all of these datasets (Supplementary Figures 3.1–3.6, Figure 3.6).

Whereas all of the sequences in the DNA-R, -S, -M, -C and -N trees fell within clearly defined SPG and AG clades (Supplementary Figures 3.1, 3.3, 3.4, 3.5 and 3.6), the DNA-U3 tree contains some sequences that cannot be convincingly classified into either the SPG or AG clades (Supplementary Figure 3.2).

Given the low degrees of support for most of the branches within the RF-CSD trees we opted to focus on the RF-FGD tree (Figure 3.6), in our assessment of finer scale geographic structure within each of the AG and SPG clades.

It is immediately evident from the RF-FGD tree that there is a high degree of geographic clustering amongst the sub-clades within both the SPG and AG. That is, there are many well supported monophyletic groups with viruses all sampled from the same country. This clustering is particularly strong amongst the SPG isolates. For example the Tongan, Hawaiian and Australian SPG isolates all form well-supported monophyletic groups that are indicative of these viruses all having originated from a single founder within their respective countries/territories, a pattern that strongly supports the occurrence of only a single major BBTV introduction event in each of these countries/territories.

In some cases, it is even possible to suggest the origins of some of the founder viruses. For example, the only two Hawaiian isolates that we examined are nested within a more diverse clade of Samoan isolates which suggests that the Hawaiian isolates may have originated from Samoa. Confirmation of this origin would require further analysis of BBTV samples from other Pacific islands around Samoa and Hawaii.

With the exception of a single divergent Egyptian isolate, all of the African BBTV isolates cluster together within a sub-clade containing viruses sampled on the Indian subcontinent. While this group of African BBTV isolates does not form a well-supported monophyletic clade, its nesting within the more diverse Indian clade suggests that the founder(s) of the Sub-Saharan African BBTV population originated on the Indian subcontinent. Although the number of times viruses were transferred to Africa from the subcontinent (i.e. the number of founding virus lineages) cannot be determined from the ML phylogenetic analysis (Figure 3.6), as with the other regions that are represented within the tree, the phylogenetic evidence is consistent with the Sub-Saharan African BBTV isolates all originating from a single introduction from India, Pakistan or Sri Lanka. It must be noted, however, that the divergent Egyptian isolate is most closely related to viruses in the AG, clearly indicating that BBTV has likely been introduced to Africa at least twice, once from the Indian subcontinent, probably into central/East Africa and once from the Asian region into North Africa.

The global patterns of geographic structure that are evident within the BBTV phylogeny are not consistent with frequent, human-mediated movements of BBTV isolates at a trans-continental-scale. They are, however, consistent with more gradual, natural or human-facilitated movements of the virus (via infected propagules and/or aphids) from its centres of diversity in India, and Southeast Asia / Far East across the banana growing regions of the world. Accordingly, there is clear evidence within the tree presented in Figure 3.6, of frequent shorter distance BBTV movements. For example, sequences sampled from Taiwan, China and Indonesia intermingle with one another within well supported clades: a pattern which suggests that there must have been multiple BBTV movements between these locations.

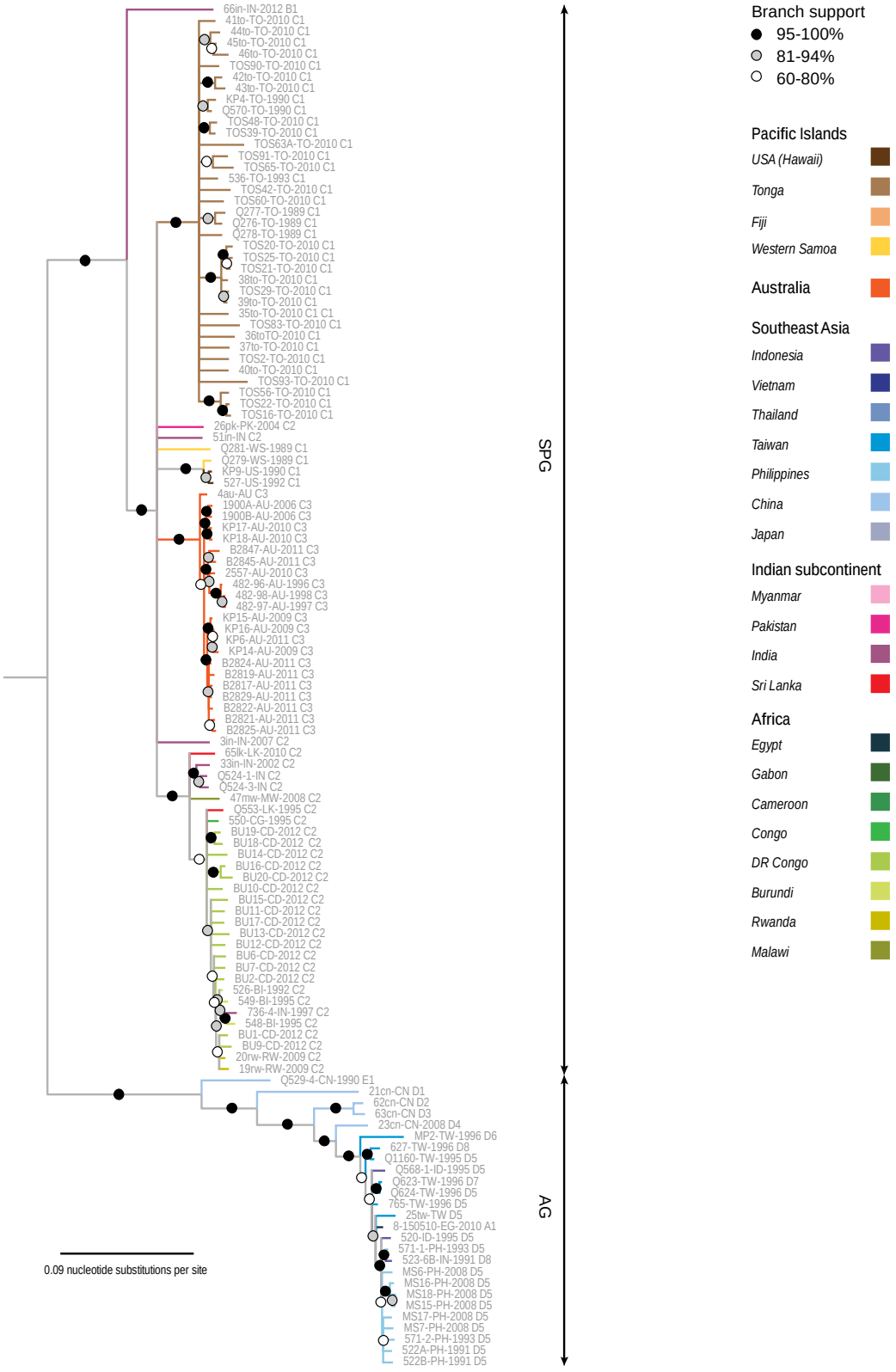


Figure 3.6: A RAxML tree of the full genome dataset (FGD) after all recombination and reassortment sequences were removed. ABTV was used to root the phylogenetic tree. Branches with <60% bootstrap support have been collapsed. Full genomes are shown with isolate name, two letter country code, year of collection and group name. GenBank accession numbers of the components which constitute each full genome can be found in Supplementary Table 3.1.

3.4.6 Selection analysis

The large amount of new genomic sequence data generated both here for BBTV and in a recent CBDV study (Savory & Ramakrishnan, 2014) presented the opportunity to, for the first time, comparatively analyse patterns of natural selection acting upon the codon sequences of different babuvirus species. Whereas differences between the selection patterns evident within the genes of two related species might reflect selective processes specific to their different niches (such as adaptation to specific host or vector species), similarities might reflect more generally relevant selective processes (such as general adaptations required for replication within and movement between plant cells). We therefore assembled codon-aligned gene datasets for all available CBDV and BBTV sequences. In addition to these babuvirus datasets we also assembled FBNYV DNA-R datasets so that we could compare selection patterns evident within *rep* genes of viruses in the two known nanovirus genera. It was not possible to directly compare selection signals in other FBNYV genes with their homologues in the babuviruses because these genes were too divergent between the two groups.

A summary of the selection acting on the coding regions of BBTV, CBDV and FBNYV is provided in Figure 3.7. As expected, all of the analysed genes in all three analysed species displayed synonymous substituting rates (dS) that were higher than their non-synonymous substitution rates (dN). Situations where gene-wide estimates of dN/dS are smaller than one are indicative of genes being well adapted to their environments such that natural selection generally disfavours change: a type of selection referred to as negative or purifying selection.

Despite the gene-wide pervasiveness of negative selection there remained strong evidence within all of the analysed BBTV and CBDV genes of selection favouring change at individual codon sites: a type of selection referred to as diversifying or positive selection. All BBTV and the majority of CBDV ORFs show both episodic and pervasive diversifying selection, DNA-C and DNA-M of CBDV are the exceptions and show only pervasive selection. FBNYV shows no positive (episodic or pervasive) selection detected in the coding regions of DNA-R, DNA-C or DNA-N. The relatively fewer sequences of FBNYV available may be contributing to the difficulty in properly detecting selection within the ORFs. BBTV shows a much larger number of sites evolving under positive selection across all ORFs than the number seen in FBNYV and CBDV.

The Rep encoded by DNA-R has six highly conserved motifs, three rolling circle replication (RCR) motifs RCR I, RCR II, RCR III and SF3 helicase Motifs, Walker A, Walker B, and Motif C which are all involved in replication (Gorbalenya *et al.*, 1990; Rosario *et al.*, 2012; Vega-Rocha *et al.*, 2007). In BBTV negative selection was detected in all DNA-R motifs except Motif C (Figure 3.7A). The only site of positive selection detected in any motifs was seen in the Walker B motif of BBTV, with episodic selection detected at a single site. Only one sequence (Isolate TOS25 [KM607693]) showed a change of an amino acid at that site with the motif CIFDI rather than VIFDI. In CBDV negative selection was detected in Walker A and Walker B and strong negative selection in RCR II. In FBNYV no selection was detected in any of the DNA-R motifs. The ORF of DNA-C encodes the Clink (Cell cycle-link) protein which contains two conserved motifs, LxCxE and F-box which bind to the plant hosts retinoblastoma binding protein (RBP) and SKP1 proteins respectively in order to gain control of the plants cell-cycle and increase viral replication (Aronson *et al.*, 2000b; Lageix *et al.*, 2007). In all three species negative selection was detected within the F-box, however, only BBTV showed negative selection in the LxCxE motif. Except in the Walker B motif all motifs showed either no selection or negative selection which, due to the conserved nature of motifs, was expected. The positive selection detected is episodic selection, so only a subset of sequences is affected, with only one sequence showing an amino acid change.

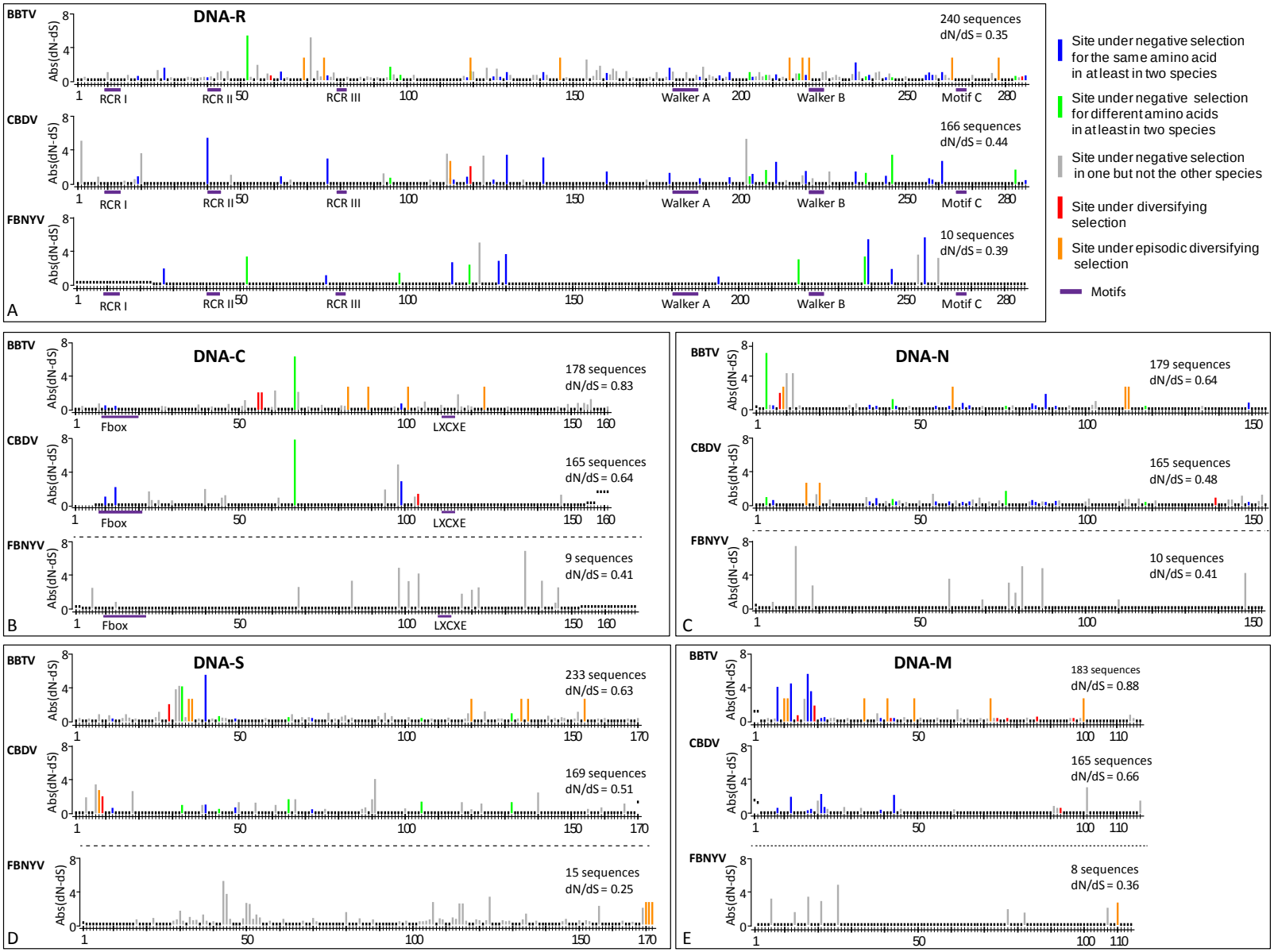


Figure 3.7: See next page for figure legend

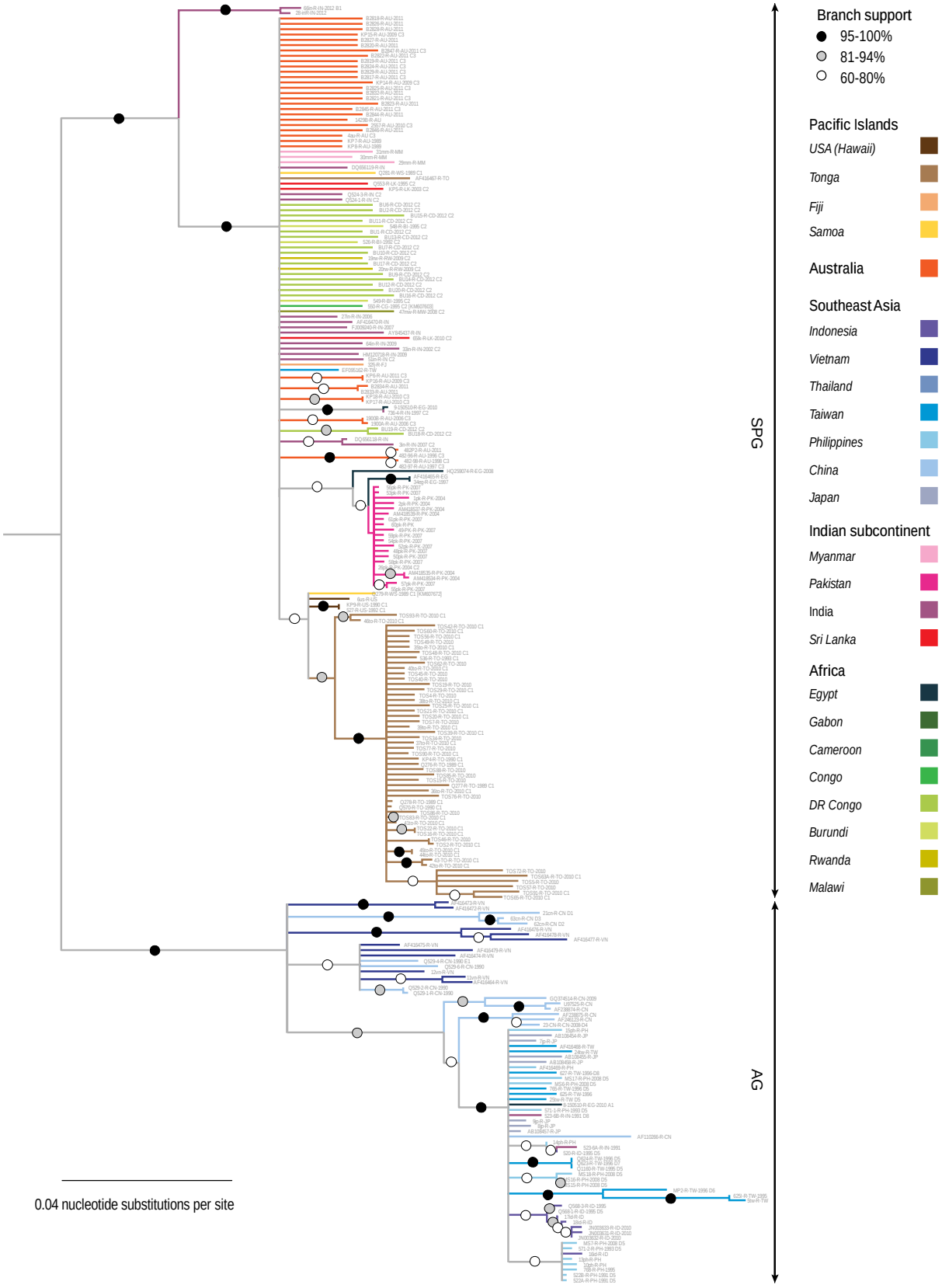
Figure 3.7: Selection analyses of the open reading frames (ORFs) of BBTV, CBDV and FBNYV. For the ORF of DNA R (A) all three species were aligned together, for all other ORFs (B-E) BBTV and CBDV were aligned together. Negative selection is shown in grey, blue and green and positive selection is shown in red and orange. Episodic selection (orange) depicts the presence of a signal at that site, for all other selection Abs (dN-dS) is shown. Motifs of DNA-R and DNA-C are shown with horizontal purple bars. Number of sequences analysed and overall dN/dS is shown for each ORF

3.5. Concluding remarks

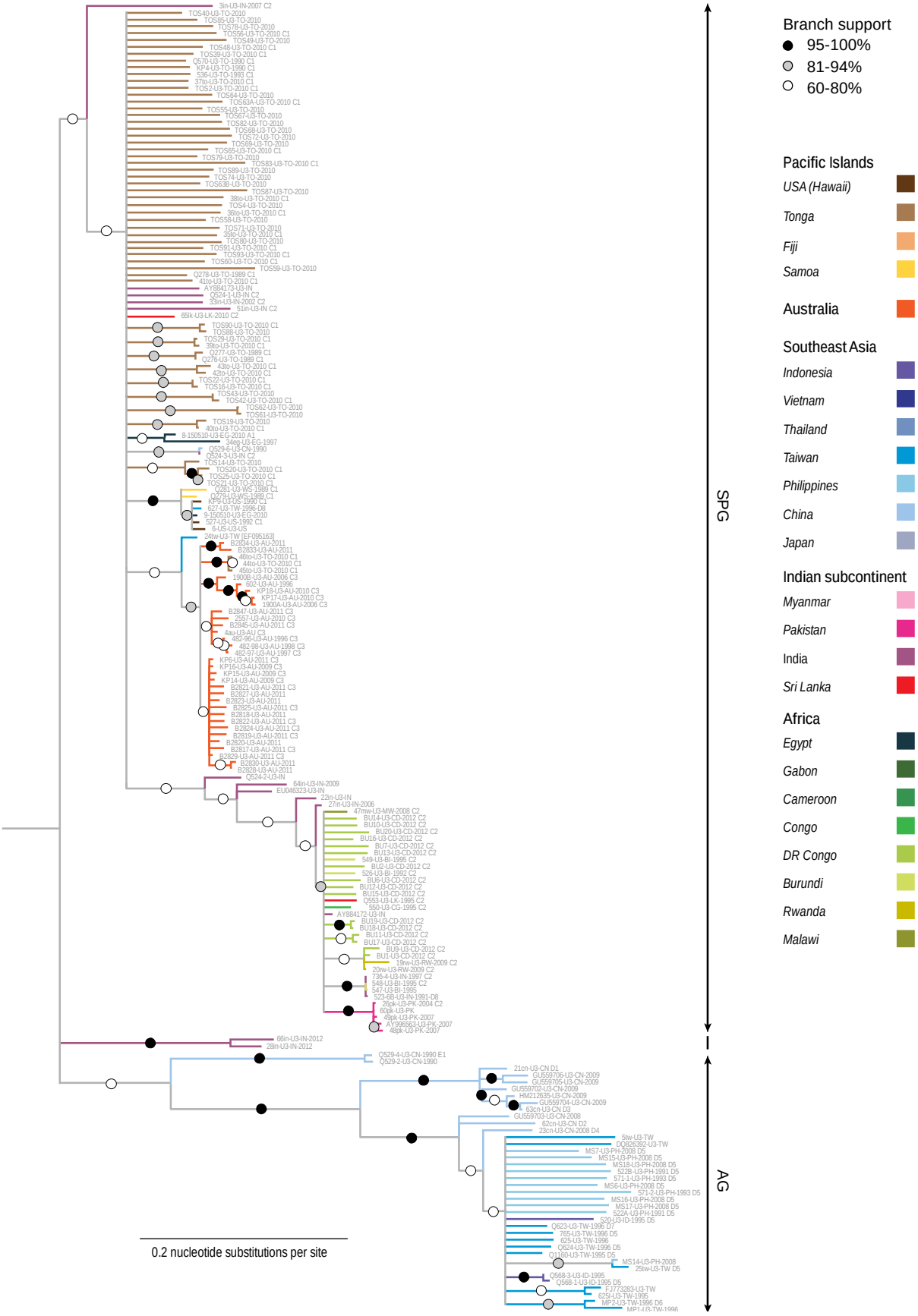
Here we have studied the landscape of global BBTV diversity to reveal that the Indian Subcontinent, Southeast Asia, and the Far East are the current BBTV diversity hotspots. Accounting for recombination and reassortment, we phylogenetically analysed 855 newly-sequenced full BBTV genome components together with all available complete genome component sequences presently available in GenBank. These analyses revealed that the global distribution of BBTV genotypes is highly structured at the continental scale, a finding that suggests human-mediated inter-continental transfers of BBTV genotypes occur relatively infrequently.

Therefore, although the geographically structured patterns of BBTV diversity that we have observed are consistent with the slow prehistoric spread of BBTV across the banana growing regions of the eastern hemisphere, it remains similarly plausible that this pattern could also have been attained with either infrequent modern human-mediated movements and/or slow geographic range expansions during the modern era.

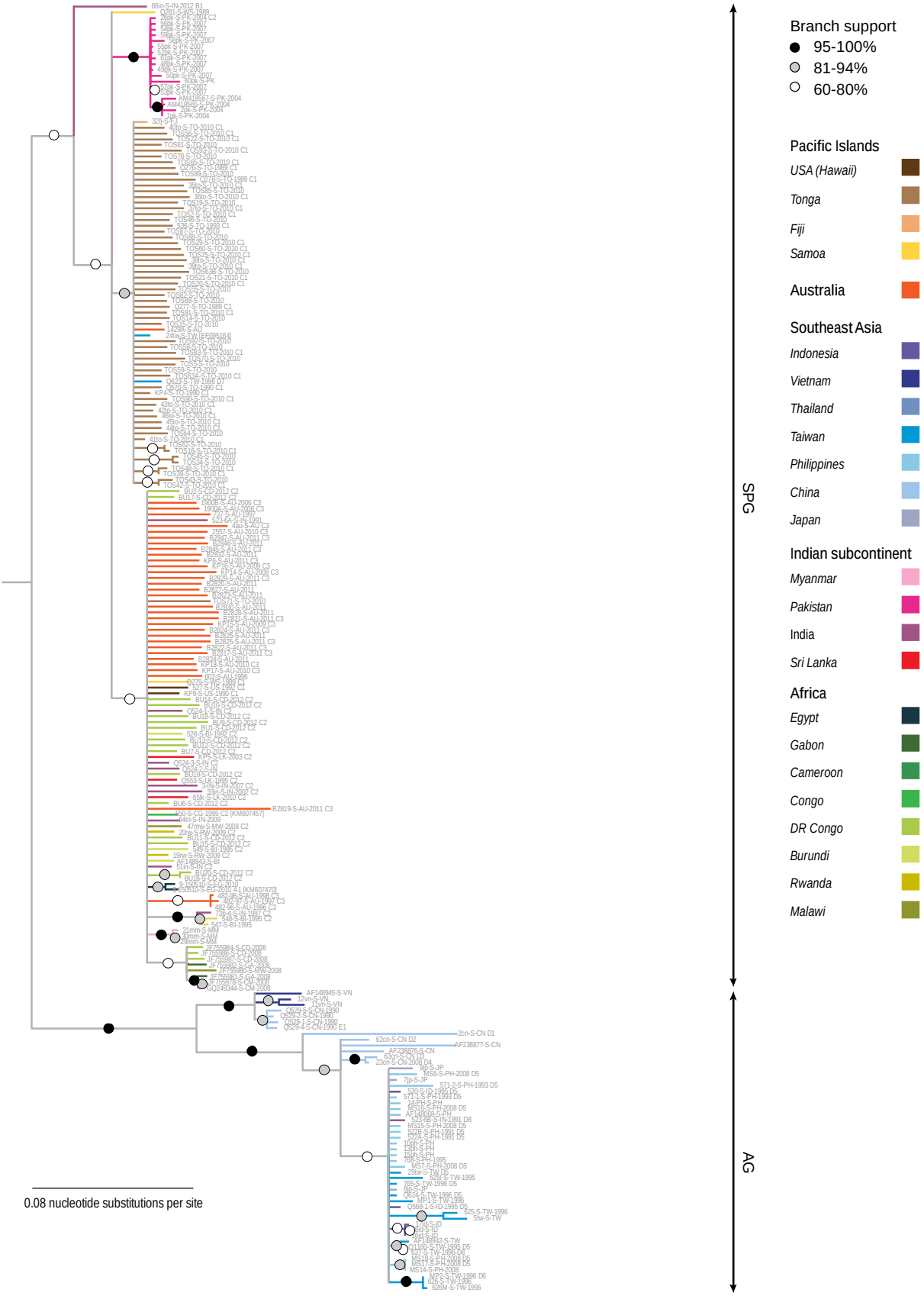
GenBank Accession number: KM607005- KM607859



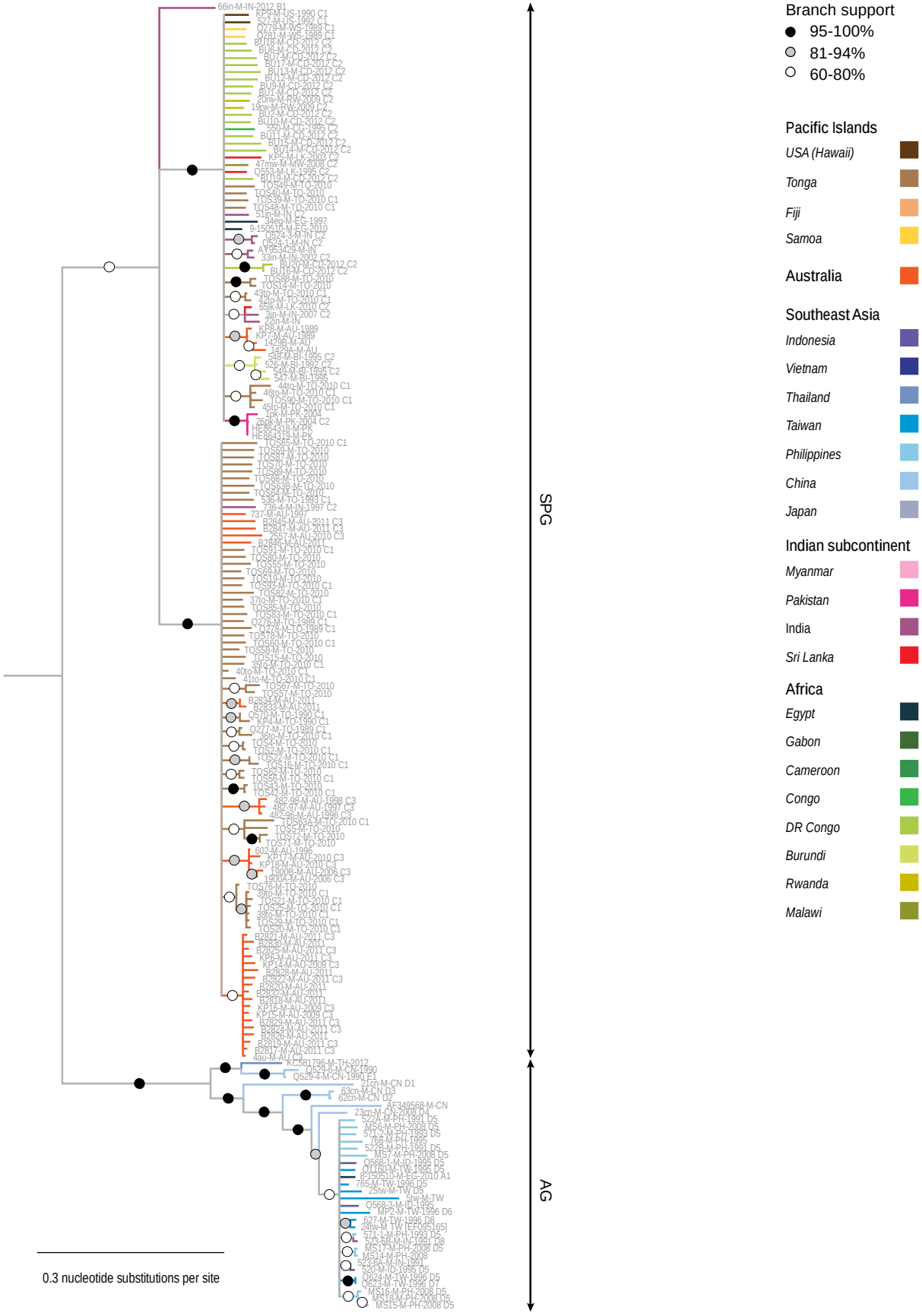
Supplementary Figure 1: Maximum likelihood phylogenetic tree constructed from the DNA-R recombination free dataset using the GTR+I+G4 nucleotide substitution model, rooted with DNA-R ABTV sequences. Branches with <60% bootstrap support have been collapsed. Study codes along with two letter country codes, years of collection and group names are given for isolates where known. GenBank accession numbers for all study codes are available in Supplementary 3.1.



Supplementary Figure 2: Maximum likelihood phylogenetic tree constructed from the BBTV DNA-U3 recombination free dataset using GTR+I+G4 nucleotide substitution model, rooted with DNA-U3 ABTV sequences. Branches with <60% bootstrap support have been collapsed. Study codes along with two letter country codes, years of collection and group names are given for isolates where known. GenBank accession numbers for all study codes are available in Supplementary 3.1.

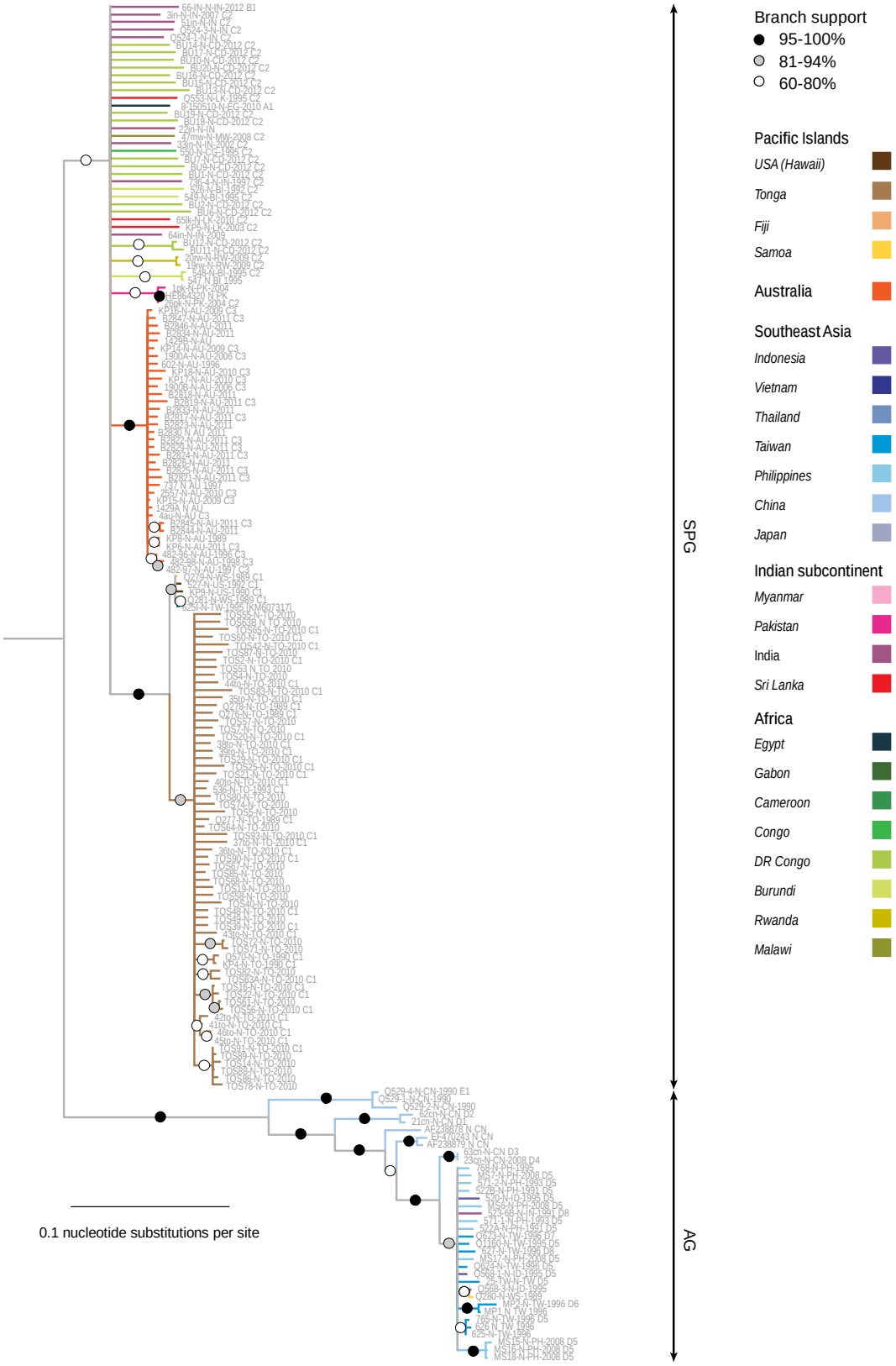


Supplementary Figure 3: Maximum likelihood phylogenetic tree constructed from the BBTV DNA-S recombination free dataset using GTR+I+G4 nucleotide substitution model, rooted with DNA-S ABTV sequences. Branches with <60% bootstrap support have been collapsed. Study codes along with two letter country codes, years of collection and group names are given for isolates where known. GenBank accession numbers for all study codes are available in Supplementary 3.1.



Supplementary Figure 4: Maximum likelihood phylogenetic tree constructed from the BBTV DNA-M recombination free dataset using GTR+I+G4 nucleotide substitution model, rooted with DNA-M ABTV sequences. Branches with <60% bootstrap support have been collapsed. Study codes along with two letter country codes, years of collection and group names are given for isolates where known. GenBank accession numbers for all study codes are available in Supplementary 3.1.

Supplementary Figure 5: Maximum likelihood phylogenetic tree constructed from the BBTV DNA-C recombination free dataset using GTR+G4 nucleotide substitution model, rooted with DNA-C ABTV sequences. Branches with <60% bootstrap support have been collapsed. Study codes along with two letter country codes, years of collection and group names are given for isolates where known. GenBank accession numbers for all study codes are available in Supplementary 3.1.



Supplementary Figure 6: Maximum likelihood phylogenetic tree constructed from the BBTV DNA-N recombination free dataset using GTR+G4 nucleotide substitution model, rooted with DNA-N ABTV sequences. Branches with <60% bootstrap support have been collapsed. Study codes along with two letter country codes, years of collection and group names are given for isolates where known. GenBank accession numbers for all study codes are available in Supplementary 3.1.

Supplementary Table: 3.1: BBTV isolate information for all sequences used in this study. In order to link components from the same sample all components are named with study code-component-country code-year-subgroup where known. Full genomes are shown in bold along with the corresponding subgroup. Accession numbers and references for all sequences are given. Countries of sample collection are identified by two letter country codes: AU-Australia, BI-Burundi, CD-Democratic Republic of Congo, CG-Congo, CM-Cameroon, CN-China, EG-Egypt, FJ-Fiji, GA-Gabon, ID-Indonesia, IN-India, JP-Japan, LK-Sri Lanka, MM-Myanmar, MW-Malawi, PH-Philippines, PK-Pakistan, RW-Rwanda, TO-Tonga, TW-Taiwan, US-United States of America [Hawaii], VN-Vietnam, WS- Samoa. ABTV sequence information is located at the bottom of the table. UP=unpublished and TS=This study

Study code	Isolate	Subgroup	Country	Sequence reference	Year	DNA-R Accession	DNA-U3 Accession	DNA-S Accession	DNA-M Accession	DNA-C accession	DNA-N Accession
737	737		AU	TS	1997			KM607467	KM607173	KM607030	KM607321
2557	**2557**	**C3**	**AU**	**TS**	**2010**	**KM607588**	**KM607724**	**KM607442**	**KM607152**	**KM607009**	**KM607298**
1429A	1429A		AU	TS				KM607439	KM607148	KM607005	KM607294
1429B	1429B		AU	TS		KM607585			KM607149	KM607006	KM607295
1900A	**1900A**	**C3**	**AU**	**TS**	**2006**	**KM607586**	**KM607722**	**KM607440**	**KM607150**	**KM607007**	**KM607296**
1900B	**1900B**	**C3**	**AU**	**TS**	**2006**	**KM607587**	**KM607723**	**KM607441**	**KM607151**	**KM607008**	**KM607297**
482-96	**482-96**	**C3**	**AU**	**TS**	**1996**	**KM607589**	**KM607725**	**KM607443**	**KM607153**	**KM607010**	**KM607299**
482-97	**482-97**	**C3**	**AU**	**TS**	**1997**	**KM607590**	**KM607726**	**KM607444**	**KM607154**	**KM607011**	**KM607300**
482-98	**482-98**	**C3**	**AU**	**TS**	**1998**	**KM607591**	**KM607727**	**KM607445**	**KM607155**	**KM607012**	**KM607301**
482P2	482P2		AU	TS	2011	KM607592					
4au		**C3**	**AU**	**(Burns *et al.*, 1995)**		**S56276**	**L41576**	**L41574**	**L41575**	**L41578**	**L41577**
B2817	**B2817**	**C3**	**AU**	**TS**	**2011**	**KM607614**	**KM607749**	**KM607472**	**KM607178**	**KM607035**	**KM607325**
B2818	B2818		AU	TS	2011	KM607615	KM607750		KM607179	KM607036	KM607326
B2819	**B2819**	**C3**	**AU**	**TS**	**2011**	**KM607616**	**KM607751**	**KM607473**	**KM607180**	**KM607037**	**KM607327**
B2820	B2820		AU	TS	2011	KM607617		KM607474	KM607181	KM607038	
B2821	**B2821**	**C3**	**AU**	**TS**	**2011**	**KM607618**	**KM607752**	**KM607475**	**KM607182**	**KM607039**	**KM607328**
B2822	**B2822**	**C3**	**AU**	**TS**	**2011**	**KM607619**	**KM607753**	**KM607476**	**KM607183**	**KM607040**	**KM607329**
B2823	B2823		AU	TS	2011	KM607620	KM607754	KM607477		KM607041	KM607330
B2824	**B2824**	**C3**	**AU**	**TS**	**2011**	**KM607621**	**KM607755**	**KM607478**	**KM607184**	**KM607042**	**KM607331**
B2825	**B2825**	**C3**	**AU**	**TS**	**2011**	**KM607622**	**KM607756**	**KM607479**	**KM607185**	**KM607043**	**KM607332**
B2826	B2826		AU	TS	2011	KM607623		KM607480	KM607186		KM607333
B2827	B2827		AU	TS	2011	KM607624	KM607757	KM607481		KM607044	
B2828	B2828		AU	TS	2011	KM607625		KM607482	KM607187	KM607045	
B2829	**B2829**	**C3**	**AU**	**TS**	**2011**	**KM607626**	**KM607758**	**KM607483**	**KM607188**	**KM607046**	**KM607334**
B2830	B2830		AU	TS	2011		KM607759	KM607484	KM607189	KM607047	KM607335
B2832	B2832		AU	TS	2011	KM607627		KM607485	KM607190	KM607048	
B2833	B2833		AU	TS	2011	KM607628	KM607760		KM607191	KM607049	KM607336
B2834	B2834		AU	TS	2011	KM607629	KM607761	KM607486	KM607192		KM607337
B2844	B2844		AU	TS	2011	KM607630				KM607050	KM607338
B2845	**B2845**	**C3**	**AU**	**TS**	**2011**	**KM607631**	**KM607762**	**KM607487**	**KM607193**	**KM607051**	**KM607339**
B2846	B2846		AU	TS	2011	KM607632		KM607488	KM607194	KM607052	KM607340
B2847	**B2847**	**C3**	**AU**	**TS**	**2011**	**KM607633**	**KM607763**	**KM607489**	**KM607195**	**KM607053**	**KM607341**
KP14	**KP14**	**C3**	**AU**	**TS**	**2009**	**KM607650**	**KM607780**	**KM607506**	**KM607212**	**KM607070**	**KM607358**
KP15	**KP15**	**C3**	**AU**	**TS**	**2009**	**KM607651**	**KM607781**	**KM607507**	**KM607213**	**KM607071**	**KM607359**
KP17	**KP17**	**C3**	**AU**	**TS**	**2010**	**KM607653**	**KM607783**	**KM607509**	**KM607215**	**KM607073**	**KM607361**
KP18	**KP18**	**C3**	**AU**	**TS**	**2010**	**KM607654**	**KM607784**	**KM607510**	**KM607216**	**KM607074**	**KM607362**
KP6	**KP6**	**C3**	**AU**	**TS**	**2011**	**KM607657**	**KM607786**	**KM607513**	**KM607219**	**KM607077**	**KM607365**
KP7	KP7		AU	TS	1989	KM607658			KM607220	KM607078	
KP8	KP8		AU	TS	1989	KM607659			KM607221	KM607079	KM607366
602	602		AU	TS	1996		KM607741	KM607460	KM607170	KM607026	KM607315
KP16	**KP16**	**C3**	**AU**	**TS**	**2009**	**KM607652**	**KM607782**	**KM607508**	**KM607214**	**KM607072**	**KM607360**
526	**526**	**C2**	**BI**	**TS**	**1992**	**KM607598**	**KM607732**	**KM607451**	**KM607161**	**KM607017**	**KM607306**
547	547		BI	TS	1995		KM607735	KM607454	KM607164	KM607020	KM607309
548	**548**	**C2**	**BI**	**TS**	**1995**	**KM607601**	**KM607736**	**KM607455**	**KM607165**	**KM607021**	**KM607310**
549	**549**	**C2**	**BI**	**TS**	**1995**	**KM607602**	**KM607737**	**KM607456**	**KM607166**	**KM607022**	**KM607311**
AF148943			BI	(Wanitchakorn *et al.*, 2000b)				AF148943			
BU1	**BU1**	**C2**	**CD**	**TS**	**2012**	**KM607634**	**KM607764**	**KM607490**	**KM607196**	**KM607054**	**KM607342**

BU10	**BU10**	**C2**	**CD**	**TS**	**2012**	**KM607635**	**KM607765**	**KM607491**	**KM607197**	**KM607055**	**KM607343**
BU11	**BU11**	**C2**	**CD**	**TS**	**2012**	**KM607636**	**KM607766**	**KM607492**	**KM607198**	**KM607056**	**KM607344**
BU12	**BU12**	**C2**	**CD**	**TS**	**2012**	**KM607637**	**KM607767**	**KM607493**	**KM607199**	**KM607057**	**KM607345**
BU13	**BU13**	**C2**	**CD**	**TS**	**2012**	**KM607638**	**KM607768**	**KM607494**	**KM607200**	**KM607058**	**KM607346**
BU14	**BU14**	**C2**	**CD**	**TS**	**2012**	**KM607639**	**KM607769**	**KM607495**	**KM607201**	**KM607059**	**KM607347**
BU15	**BU15**	**C2**	**CD**	**TS**	**2012**	**KM607640**	**KM607770**	**KM607496**	**KM607202**	**KM607060**	**KM607348**
BU16	**BU16**	**C2**	**CD**	**TS**	**2012**	**KM607641**	**KM607771**	**KM607497**	**KM607203**	**KM607061**	**KM607349**
BU17	**BU17**	**C2**	**CD**	**TS**	**2012**	**KM607642**	**KM607772**	**KM607498**	**KM607204**	**KM607062**	**KM607350**
BU18	**BU18**	**C2**	**CD**	**TS**	**2012**	**KM607643**	**KM607773**	**KM607499**	**KM607205**	**KM607063**	**KM607351**
BU19	**BU19**	**C2**	**CD**	**TS**	**2012**	**KM607644**	**KM607774**	**KM607500**	**KM607206**	**KM607064**	**KM607352**
BU2	**BU2**	**C2**	**CD**	**TS**	**2012**	**KM607645**	**KM607775**	**KM607501**	**KM607207**	**KM607065**	**KM607353**
BU20	**BU20**	**C2**	**CD**	**TS**	**2012**	**KM607646**	**KM607776**	**KM607502**	**KM607208**	**KM607066**	**KM607354**
BU6	**BU6**	**C2**	**CD**	**TS**	**2012**	**KM607647**	**KM607777**	**KM607503**	**KM607209**	**KM607067**	**KM607355**
BU7	**BU7**	**C2**	**CD**	**TS**	**2012**	**KM607648**	**KM607778**	**KM607504**	**KM607210**	**KM607068**	**KM607356**
BU9	**BU9**	**C2**	**CD**	**TS**	**2012**	**KM607649**	**KM607779**	**KM607505**	**KM607211**	**KM607069**	**KM607357**
JF755984	DRC-TV24.9		CD	(Kumar *et al.*, 2011)	2008			JF755984			
JF755986	DRC-23.3		CD	(Kumar *et al.*, 2011)	2008			JF755986			
JF755987	DRC-25.2		CD	(Kumar *et al.*, 2011)	2008			JF755987			
550	**550**	**C2**	**CG**	**TS**	**1995**	**KM607603**	**KM607738**	**KM607457**	**KM607167**	**KM607023**	**KM607312**
GQ249344			CM	UP	2008			GQ249344			
JF755978	CAM-TV4.1		CM	(Kumar *et al.*, 2011)	2008			JF755978			
21cn	**Hainan**	**D1**	**CN**	**(Jun & Zhi-Xin, 2005)**		**AY450396**	**AY606084**	**AY494786**	**AY494788**	**AY606085**	**AY494787**
23cn	**Haikou**	**D4**	**CN**	**UP**	**2008**	**FJ463042**	**FJ463043**	**FJ463044**	**FJ463045**	**FJ463046**	**FJ463047**
62cn	**Haikou 4**	**D2**	**CN**	**(Yu *et al.*, 2012)**		**HQ378190**	**HM231314**	**HQ378191**	**HQ378192**	**HQ378193**	**HQ378194**
63cn	**Haikou 2**	**D3**	**CN**	**(Yu *et al.*, 2012)**		**HQ616074**	**HQ616075**	**HQ616076**	**HQ616077**	**HQ616078**	**HQ616079**
AF110266	Zhangzhou		CN	(La *et al.*, 2000)		AF110266					
AF238874	NS strain		CN	(He *et al.*, 2000)		AF238874					
AF238875	NSP strain		CN	(He *et al.*, 2000)		AF238875					
AF238876	NS strain		CN	(He *et al.*, 2001a)				AF238876			
AF238877	NSP strain		CN	(He *et al.*, 2001a)				AF238877			
AF238878	NS strain		CN	(He *et al.*, 2001b)							AF238878
AF238879	NSP strain		CN	(He *et al.*, 2001b)							AF238879
AF246123	Guangdong-1		CN	UP		AF246123					
AF349568	Zhangzhou		CN	UP					AF349568		
AY264347	NSP strain		CN	UP						AY264347	
AY266417	NS strain		CN	UP						AY266417	
EF470243	NSP strain		CN	UP							EF470243
GQ374514	Zhangjiang		CN	UP	2009	GQ374514					
GU559702	ChengMai		CN	UP	2009		GU559702				
GU559703	LeDong		CN	UP	2008		GU559703				
GU559704	Haikou 1		CN	UP	2009		GU559704				
GU559705	DanZhou		CN	UP	2009		GU559705				
GU559706	DanZhouHD		CN	UP	2009		GU559706				
HM212635	Haikou 3		CN	UP	2009		HM212635				
Q529-1	Q529-1		CN	TS	1990	KM607676		KM607533			KM607385
Q529-2	Q529-2		CN	TS	1990	KM607677	KM607806	KM607534		KM607098	KM607386
Q529-4	**Q529-4**	**E1**	**CN**	**TS**	**1990**	**KM607678**	**KM607807**	**KM607535**	**KM607239**	**KM607099**	**KM607387**
Q529-5	Q529-5		CN	TS	1990			KM607536			
Q529-6	Q529-6		CN	TS	1990	KM607679	KM607808		KM607240		
U97525	C4		CN	UP		U97525					
34eg	Kalubia		EG	UP	1997	AF102780	AF102781		AF102783		
8-150510	**8-150510**	**A1**	**EG**	**TS**	**2010**	**KM607612**	**KM607747**	**KM607470**	**KM607176**	**KM607033**	**KM607324**
9-150510	9-150510		EG	TS	2010	KM607613	KM607748	KM607471	KM607177	KM607034	
AF102784			EG	UP	1997					AF102784	
AF416465			EG	(Karan *et al.*, 1994)		AF416465					
HQ259074			EG	(Abdel-Salam *et al.*, 2012)	2008	HQ259074					
32fj			FJ	(Wanitchakorn *et al.*, 2000b), (Karan *et al.*, 1994)		AF416466		AF148944			

JF755981	GAB-TV18.2		GA	(Kumar *et al.*, 2011)	2008			JF755981			
JF755982	GAB-TV-17.5		GA	(Kumar *et al.*, 2011)	2008			JF755982			
520	**520**	**D5**	**ID**	**TS**	**1995**	**KM607593**	**KM607728**	**KM607446**	**KM607156**	**KM607013**	**KM607302**
16id	IG33		ID	(Furuya *et al.*, 2004)	1998-2000	AB186924		AB186927			
17id	IG64		ID	(Furuya *et al.*, 2004)	1998-2000	AB186925		AB186928			
18id	IJs11		ID	(Furuya *et al.*, 2004)	1998-2000	AB186926		AB186929			
JN003631	Bali Tukad Badung II		ID	(Pinili *et al.*, 2011)	2010	JN003631					
JN003632	Bali Tukad Petanu I		ID	(Pinili *et al.*, 2011)	2010	JN003632					
JN003633	Bali Sempidi I		ID	(Pinili *et al.*, 2011)	2010	JN003633					
Q568-1	**Q568-1**	**D5**	**ID**	**TS**	**1995**	**KM607681**	**KM607810**	**KM607538**	**KM607242**	**KM607101**	**KM607389**
Q568-3	Q568-3		ID	TS	1995	KM607682	KM607811		KM607243	KM607102	KM607390
22in	BT-1		IN	UP			AY960129		AY948439		AY948438
27in	Kerala 1		IN	UP	2006	FJ009238	FJ009239				
28in	Meghalaya		IN	UP	2012	JQ911667	JQ911668				
33in	**HB-TN**	**C2**	**IN**	**(Selvarajan *et al.*, 2010)**	**2002**	**EU140342**	**EU140341**	**EU589459**	**EU190971**	**EU190969**	**EU190970**
3in	**Lucknow**	**C2**	**IN**	**(Vishnoi *et al.*, 2009)**	**2006-2007**	**DQ256267**	**EU402601**	**EF687856**	**EU516323**	**EU051379**	**EU391633**
51in	**Bihar**	**C2**	**IN**	**(Islam *et al.*, 2010a)**		**FJ605506**	**FJ605508**	**FJ605507**	**FJ609642**	**FJ609643**	**FJ609644**
523-6A	523-6A		IN	TS	1991	KM607596		KM607449	KM607159		
523-6B	**523-6B**	**D8**	**IN**	**TS**	**1991**	**KM607597**	**KM607731**	**KM607450**	**KM607160**	**KM607016**	**KM607305**
64in	Bangalore-GKVK		IN	UP	2009	JN243751	JN243752	JN243753			JN243754
66in	**Umiam**	**B1**	**IN**	**(Banerjee *et al.*, 2014)**	**2012**	**KC119098**	**KC466373**	**KC466374**	**KC466375**	**KC466376**	**KC466377**
736-4	**736-4**	**C2**	**IN**	**TS**	**1997**	**KM607609**	**KM607745**	**KM607466**	**KM607172**	**KM607029**	**KM607320**
AF416470			IN	(Karan *et al.*, 1994)		AF416470					
AY845437			IN	UP		AY845437					
AY884172	BBTR1		IN	(Anandhi *et al.*, 2007)			AY884172				
AY884173	BRJT9		IN	(Anandhi *et al.*, 2007)			AY884173				
AY953429	TN		IN	UP					AY953429		
DQ656118	Kanpur		IN	UP		DQ656118					
DQ656119	Etawah		IN	UP		DQ656119					
EU046323	Bangalore		IN	UP			EU046323				
FJ009240	Kerala 2		IN	UP	2007	FJ009240					
HM120718	Delhi		IN	UP	2009	HM120718					
Q524-1	**Q524-1**	**C2**	**IN**	**TS**		**KM607674**	**KM607803**	**KM607530**	**KM607237**	**KM607095**	**KM607383**
Q524-2	Q524-2		IN	TS			KM607804	KM607531		KM607096	
Q524-3	**Q524-3**	**C2**	**IN**	**TS**		**KM607675**	**KM607805**	**KM607532**	**KM607238**	**KM607097**	**KM607384**
7jp	JN4		JP	(Furuya *et al.*, 2005)	1999-2003	AB108452		AB108449			
8jp	JK3		JP	(Furuya *et al.*, 2005)	1999-2003	AB108453		AB108450			
9jp	JY1		JP	(Furuya *et al.*, 2005)	1999-2003	AB108456		AB108451			
AB108454	JM5		JP	(Furuya *et al.*, 2005)	1999-2003	AB108454					
AB108455	JM6		JP	(Furuya *et al.*, 2005)	1999-2003	AB108455					
AB108457	JY3		JP	(Furuya *et al.*, 2005)	1999-2003	AB108457					
AB108458	JY7		JP	(Furuya *et al.*, 2005)	1999-2003	AB108458					
65lk	**Kandy**	**C2**	**LK**	**UP**	**2010**	**JN250593**	**JN250594**	**JN250595**	**JN250596**	**JN250597**	**JN250598**
KP5	KP5		LK	TS	2003	KM607656		KM607512	KM607218	KM607076	KM607364
Q553	**Q553**	**C2**	**LK**	**TS**	**1995**	**KM607680**	**KM607809**	**KM607537**	**KM607241**	**KM607100**	**KM607388**
29mm	My01		MM	UP		AB252639		AB252642			
30mm	MY02		MM	UP		AB252640		AB252643			
31mm	MY03		MM	UP		AB252641		AB252644			

JF755980	MAL-TV5.4		MW	(Kumar *et al.*, 2011)	2008			JF755980			
47mw	**Malawi 73**	**C2**	**MW**	**(James, 2011)**	**2008**	**JQ820453**	**JQ820454**	**JQ820455**	**JQ820456**	**JQ820457**	**JQ820458**
768	768		PH	TS	1995	KM607611		KM607469	KM607175	KM607032	KM607323
10ph	bP5		PH	(Furuya *et al.*, 2005)		AB189067		AB189068			
13ph	aP32		PH	(Furuya *et al.*, 2006)	1999-2003	AB250953		AB250956			
14ph	aP34		PH	(Furuya *et al.*, 2006)	1999-2003	AB250954		AB250957			
15ph	bP34		PH	(Furuya *et al.*, 2006)	1999-2003	AB250955		AB250958			
522A	**522A**	**D5**	**PH**	**TS**	**1991**	**KM607594**	**KM607729**	**KM607447**	**KM607157**	**KM607014**	**KM607303**
522B	**522B**	**D5**	**PH**	**TS**	**1991**	**KM607595**	**KM607730**	**KM607448**	**KM607158**	**KM607015**	**KM607304**
571-1	**571-1**	**D5**	**PH**	**TS**	**1993**	**KM607604**	**KM607740**	**KM607459**	**KM607169**	**KM607025**	**KM607314**
571-2	**571-2**	**D5**	**PH**	**TS**	**1993**	**KM607605**	**KM607739**	**KM607458**	**KM607168**	**KM607024**	**KM607313**
AF148068			PH	(Wanitchakorn *et al.*, 2000b)				AF148068			
AF416469			PH	(Karan *et al.*, 1994)		AF416469					
MS14	MS14		PH	TS	2008		KM607790	KM607517	KM607224	KM607082	
MS15	**MS15**	**D5**	**PH**	**TS**	**2008**	**KM607662**	**KM607791**	**KM607518**	**KM607225**	**KM607083**	**KM607370**
MS16	**MS16**	**D5**	**PH**	**TS**	**2008**	**KM607663**	**KM607792**	**KM607519**	**KM607226**	**KM607084**	**KM607371**
MS17	**MS17**	**D5**	**PH**	**TS**	**2008**	**KM607664**	**KM607793**	**KM607520**	**KM607227**	**KM607085**	**KM607372**
MS18	**MS18**	**D5**	**PH**	**TS**	**2008**	**KM607665**	**KM607794**	**KM607521**	**KM607228**	**KM607086**	**KM607373**
MS6	**MS6**	**D5**	**PH**	**TS**	**2008**	**KM607666**	**KM607795**	**KM607522**	**KM607229**	**KM607087**	**KM607374**
MS7	**MS7**	**D5**	**PH**	**TS**	**2008**	**KM607667**	**KM607796**	**KM607523**	**KM607230**	**KM607088**	**KM607375**
1pk	Thatha		PK	(Amin *et al.*, 2008)	2004	AM418538		AM418566	AM418541	AM418569	AM418568
26pk	**TJ1**	**C2**	**PK**	**(Hyder, 2009), (Hyder *et al.*, 2007), (Hyder *et al.*, 2011)**	**2004**	**AY996562**	**GQ214699**	**EF593169**	**EU095948**	**EF520722**	**EF529519**
2pk	Tandojam		PK	(Amin *et al.*, 2008)	2004	AM418536		AM418540		AM418564	
48pk	GH1		PK	(Hyder, 2009), (Hyder *et al.*, 2011)	2007	FJ859722	FJ859748	FJ859735			
49pk	JS1		PK	(Hyder, 2009), (Hyder *et al.*, 2011)	2007	FJ859732	FJ859749	FJ859745			
50pk	KP1		PK	(Hyder, 2009)	2007	FJ859723		FJ859736			
52pk	KP2		PK	(Hyder, 2009)	2007	FJ859724		FJ859737			
53pk	TA1		PK	(Hyder, 2009)	2007	FJ859725		Fj859738			
54pk	TA2		PK	(Hyder, 2009)	2007	FJ859726		FJ859739			
55pk	BS1		PK	(Hyder, 2009)	2007	FJ859727		FJ859740			
56pk	BS2		PK	(Hyder, 2009)	2007	FJ859728		FJ859741			
57pk	MT1		PK	(Hyder, 2009)	2007	FJ859729		FJ859742			
58pk	MT2		PK	(Hyder, 2009)	2007	FJ859730		FJ859743			
59pk	NS1		PK	(Hyder, 2009)	2007	FJ859731		FJ859744			
60pk	HD1		PK	(Hyder, 2009), (Hyder *et al.*, 2011)	2007	FJ859733	FJ859750	FJ859746			
61pk	HD2		PK	(Hyder, 2009)	2007	FJ859734		FJ859747			
AM418534	Nawabshah		PK	(Amin *et al.*, 2008)	2004	AM418534					
AM418535	Sakrand		PK	(Amin *et al.*, 2008)	2004	AM418535					
AM418537	Chambar		PK	(Amin *et al.*, 2008)	2004	AM418537					
AM418539	Hala		PK	(Amin *et al.*, 2008)	2004	AM418539					
AM418565	Kisanamari		PK	(Amin *et al.*, 2008)	2004			AM418565			
AM418567	Nasarpur		PK	(Amin *et al.*, 2008)	2004			AM418567			
AY996563	KHI		PK	(Hyder *et al.*, 2011)	2007		AY996563				
HE864318			PK	(Bashir *et al.*, 2012)					HE864318		

HE864319			PK	(Bashir *et al.*, 2012)					HE864319		
HE864320			PK	(Bashir *et al.*, 2012)							HE864320
19rw	**Rwanda 138**	**C2**	**RW**	**(James, 2011)**	**2009**	**JQ820459**	**JQ820460**	**JQ820461**	**JQ820462**	**JQ820463**	**JQ820464**
20rw	**Rwanda 142**	**C2**	**RW**	**(James, 2011)**	**2009**	**JQ820465**	**JQ820466**	**JQ820467**	**JQ820468**	**JQ820469**	**JQ820470**
KC581796	Ubonratchathanee		TH	UP	2012				KC581796		
536	**536**	**C1**	**TO**	**TS**	**1993**	**KM607600**	**KM607734**	**KM607453**	**KM607163**	**KM607019**	**KM607308**
35to	**TO166**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957628**	**JF957640**	**JF957652**	**JF957664**	**JF957676**	**JF957688**
36to	**TO114**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957625**	**JF957637**	**JF957649**	**JF957661**	**JF957673**	**JF957685**
37to	**TO121**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957626**	**JF957638**	**JF957650**	**JF957662**	**JF957674**	**JF957686**
38to	**TOS28**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957636**	**JF957648**	**JF957660**	**JF957672**	**JF957684**	**JF957696**
39to	**TOS12**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957635**	**JF957647**	**JF957659**	**JF957671**	**JF957683**	**JF957695**
40to	**TO124**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957627**	**JF957639**	**JF957651**	**JF957663**	**JF957675**	**JF957687**
41to	**TO306**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957632**	**JF957644**	**JF957656**	**JF957668**	**JF957680**	**JF957692**
42to	**TO314**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957634**	**JF957646**	**JF957658**	**JF957670**	**JF957682**	**JF957694**
43to	**TO310**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957633**	**JF957645**	**JF957657**	**JF957669**	**JF957681**	**JF957693**
44to	**TO208**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957629**	**JF957641**	**JF957653**	**JF957665**	**JF957677**	**JF957689**
45to	**TO224**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957630**	**JF957642**	**JF957654**	**JF957666**	**JF957678**	**JF957690**
46to	**TO290**	**C1**	**TO**	**(Stainton *et al.*, 2012)**	**2010**	**JF957631**	**JF957643**	**JF957655**	**JF957667**	**JF957679**	**JF957691**
AF416467			TO	(Karan *et al.*, 1994)		AF416467					
KP4	**KP4**	**C1**	**TO**	**TS**	**1990**	**KM607655**	**KM607785**	**KM607511**	**KM607217**	**KM607075**	**KM607363**
Q276	**Q276**	**C1**	**TO**	**TS**	**1989**	**KM607669**	**KM607798**	**KM607525**	**KM607232**	**KM607090**	**KM607377**
Q277	**Q277**	**C1**	**TO**	**TS**	**1989**	**KM607670**	**KM607799**	**KM607526**	**KM607233**	**KM607091**	**KM607378**
Q278	**Q278**	**C1**	**TO**	**TS**	**1989**	**KM607671**	**KM607800**	**KM607527**	**KM607234**	**KM607092**	**KM607379**
Q570	**Q570**	**C1**	**TO**	**TS**	**1990**	**KM607683**	**KM607812**	**KM607539**	**KM607244**	**KM607103**	**KM607391**
TOS14	TOS14		TO	TS	2010		KM607815	KM607542	KM607247	KM607106	KM607394
TOS15	TOS15		TO	TS	2010	KM607686		KM607543	KM607248	KM607107	
TOS16	**TOS16**	**C1**	**TO**	**TS**	**2010**	**KM607687**	**KM607816**	**KM607544**	**KM607249**	**KM607108**	**KM607395**
TOS19	TOS19		TO	TS	2010	KM607688	KM607817	KM607545	KM607250		KM607396
TOS2	**TOS2**	**C1**	**TO**	**TS**	**2010**	**KM607689**	**KM607818**	**KM607546**	**KM607251**	**KM607109**	**KM607397**
TOS20	**TOS20**	**C1**	**TO**	**TS**	**2010**	**KM607690**	**KM607819**	**KM607547**	**KM607252**	**KM607110**	**KM607398**
TOS21	**TOS21**	**C1**	**TO**	**TS**	**2010**	**KM607691**	**KM607820**	**KM607548**	**KM607253**	**KM607111**	**KM607399**
TOS22	**TOS22**	**C1**	**TO**	**TS**	**2010**	**KM607692**	**KM607821**	**KM607549**	**KM607254**	**KM607112**	**KM607400**
TOS25	**TOS25**	**C1**	**TO**	**TS**	**2010**	**KM607693**	**KM607822**	**KM607550**	**KM607255**	**KM607113**	**KM607401**
TOS29	**TOS29**	**C1**	**TO**	**TS**	**2010**	**KM607694**	**KM607823**	**KM607551**	**KM607256**	**KM607114**	**KM607402**
TOS34	TOS34		TO	TS	2010	KM607695		KM607552			
TOS39	**TOS39**	**C1**	**TO**	**TS**	**2010**	**KM607696**	**KM607824**	**KM607553**	**KM607257**	**KM607115**	**KM607403**
TOS4	TOS4		TO	TS	2010	KM607697	KM607825		KM607258	KM607116	KM607404
TOS40	TOS40		TO	TS	2010	KM607698	KM607826		KM607259	KM607117	KM607405
TOS42	**TOS42**	**C1**	**TO**	**TS**	**2010**	**KM607699**	**KM607827**	**KM607554**	**KM607260**	**KM607118**	**KM607406**
TOS43	TOS43		TO	TS	2010		KM607828	KM607555	KM607261	KM607119	
TOS45	TOS45		TO	TS	2010	KM607700		KM607556			
TOS46	TOS46		TO	TS	2010	KM607701		KM607557			
TOS48	**TOS48**	**C1**	**TO**	**TS**	**2010**	**KM607702**	**KM607829**	**KM607558**	**KM607262**	**KM607120**	**KM607407**
TOS49	TOS49		TO	TS	2010	KM607703	KM607830		KM607263	KM607121	KM607408
TOS5	TOS5		TO	TS	2010	KM607704		KM607559	KM607264		KM607409
TOS53	TOS53		TO	TS	2010						KM607410
TOS55	TOS55		TO	TS	2010		KM607831	KM607560	KM607265	KM607122	KM607411
TOS56	**TOS56**	**C1**	**TO**	**TS**	**2010**	**KM607705**	**KM607832**	**KM607561**	**KM607266**	**KM607123**	**KM607412**
TOS57	TOS57		TO	TS	2010	KM607706			KM607267	KM607124	KM607413
TOS58	TOS58		TO	TS	2010		KM607833	KM607562	KM607268	KM607125	KM607414
TOS59	TOS59		TO	TS	2010		KM607834	KM607563	KM607269	KM607126	
TOS60	**TOS60**	**C1**	**TO**	**TS**	**2010**	**KM607707**	**KM607835**	**KM607564**	**KM607270**	**KM607127**	**KM607415**
TOS61	TOS61		TO	TS	2010		KM607836	KM607565		KM607128	KM607416

TOS62	TOS62		TO	TS	2010	KM607708	KM607837	KM607566	KM607271	KM607129		
TOS63A	**TOS63A**	**C1**	**TO**	**TS**	**2010**	**KM607709**	**KM607838**	**KM607567**	**KM607272**	**KM607130**	**KM607417**	
TOS63B	TOS63B		TO	TS	2010		KM607839	KM607568	KM607273	KM607131	KM607418	
TOS64	TOS64		TO	TS	2010		KM607840	KM607569	KM607274	KM607132	KM607419	
TOS65	**TOS65**	**C1**	**TO**	**TS**	**2010**	**KM607710**	**KM607841**	**KM607570**	**KM607275**	**KM607133**	**KM607420**	
TOS67	TOS67		TO	TS	2010		KM607842		KM607276	KM607134	KM607421	
TOS68	TOS68		TO	TS	2010		KM607843	KM607571	KM607277	KM607135	KM607422	
TOS69	TOS69		TO	TS	2010		KM607844		KM607278	KM607136		
TOS7	TOS7		TO	TS	2010	KM607711					KM607423	
TOS70	TOS70		TO	TS	2010			KM607572	KM607279	KM607137		
TOS71	TOS71		TO	TS	2010		KM607845	KM607573	KM607280	KM607138	KM607424	
TOS72	TOS72		TO	TS	2010	KM607712	KM607846		KM607281	KM607139	KM607425	
TOS74	TOS74		TO	TS	2010		KM607847				KM607426	
TOS76	TOS76		TO	TS	2010	KM607713			KM607282	KM607140		
TOS77	TOS77		TO	TS	2010	KM607714						
TOS78	TOS78		TO	TS	2010		KM607848	KM607574	KM607283	KM607141	KM607427	
TOS79	TOS79		TO	TS	2010		KM607849					
TOS80	TOS80		TO	TS	2010		KM607850		KM607284		KM607428	
TOS82	TOS82		TO	TS	2010		KM607851	KM607575	KM607285	KM607142	KM607429	
TOS83	**TOS83**	**C1**	**TO**	**TS**	**2010**	**KM607715**	**KM607852**	**KM607576**	**KM607286**	**KM607143**	**KM607430**	
TOS85	TOS85		TO	TS	2010	KM607716	KM607853	KM607577	KM607287		KM607431	
TOS86	TOS86		TO	TS	2010	KM607717					KM607432	
TOS87	TOS87		TO	TS	2010		KM607854	KM607578	KM607288	KM607144	KM607433	
TOS88	TOS88		TO	TS	2010	KM607718	KM607855	KM607579	KM607289		KM607434	
TOS89	TOS89		TO	TS	2010		KM607856	KM607580	KM607290		KM607435	
TOS90	**TOS90**	**C1**	**TO**	**TS**	**2010**	**KM607719**	**KM607857**	**KM607581**	**KM607291**	**KM607145**	**KM607436**	
TOS91	**TOS91**	**C1**	**TO**	**TS**	**2010**	**KM607720**	**KM607858**	**KM607582**	**KM607292**	**KM607146**	**KM607437**	
TOS92	TOS92		TO	TS	2010			KM607583				
TOS93	**TOS93**	**C1**	**TO**	**TS**	**2010**	**KM607721**	**KM607859**	**KM607584**	**KM607293**	**KM607147**	**KM607438**	
625	625		TW	TS	1996	KM607606	KM607742	KM607461		KM607027	KM607316	
626	626		TW	TS	1996			KM607463			KM607318	
627	**627**	**D8**	**TW**	**TS**	**1996**	**KM607608**	**KM607744**	**KM607465**	**KM607171**	**KM607028**	**KM607319**	
765	**765**	**D5**	**TW**	**TS**	**1996**	**KM607610**	**KM607746**	**KM607468**	**KM607174**	**KM607031**	**KM607322**	
24tw	V-1		TW	(Hu *et al.*, 2007)		EF095161	EF095163	EF095164	EF095165	EF095166		
25tw	**Severe phenotype**	**D5**	**TW**	**(Hu *et al.*, 2007)**		**DQ826390**	**DQ826391**	**DQ826393**	**DQ826394**	**DQ826395**	**DQ826396**	
5tw	TW3		TW	(Fu *et al.*, 2009)		EU366169	EU366170	EU366171	EU366172	EU366173		
625I	625I		TW	TS	1995	KM607607	KM607743	KM607462			KM607317	
626M	626M		TW	TS	1995			KM607464				
AF148942			TW	(Wanitchakorn *et al.*, 2000b)				AF148942				
AF416468			TW	(Karan *et al.*, 1994)		AF416468						
DQ826392	Severe strain clone b		TW	(Hu *et al.*, 2007)			DQ826392					
EF095162	V-1 clone a		TW	(Hu *et al.*, 2007)		EF095162						
FJ773283	TW3 clone a		TW	(Fu *et al.*, 2009)			FJ773283					
MP1	MP1		TW	TS	1996		KM607788	KM607515			KM607368	
MP2	**MP2**	**D6**	**TW**	**TS**	**1996**	**KM607661**	**KM607789**	**KM607516**	**KM607223**	**KM607081**	**KM607369**	
Q1160	**Q1160**	**D5**	**TW**	**TS**	**1995**	**KM607668**	**KM607797**	**KM607524**	**KM607231**	**KM607089**	**KM607376**	
Q623	**Q623**	**D7**	**TW**	**TS**	**1996**	**KM607684**	**KM607813**	**KM607540**	**KM607245**	**KM607104**	**KM607392**	
Q624	**Q624**	**D5**	**TW**	**TS**	**1996**	**KM607685**	**KM607814**	**KM607541**	**KM607246**	**KM607105**	**KM607393**	
527	**527**	**C1**	**US**	**TS**	**1992**	**KM607599**	**KM607733**	**KM607452**	**KM607162**	**KM607018**	**KM607307**	
6us	Hawaiian		US	(Xie & Hu, 1995)		U18077	U18078			U18079		
KP9	**KP9**	**C1**	**US**	**TS**	**1990**	**KM607660**	**KM607787**	**KM607514**	**KM607222**	**KM607080**	**KM607367**	
11vn	Isolate V6		VN	(Furuya *et al.*, 2005)		AB113659		AB113661				
12vn	Isolate V14		VN	(Furuya *et al.*, 2005)		AB113660		AB113662				
AF148945			VN	(Wanitchakorn *et al.*, 2000b)				AF148945				
AF416464			VN	(Karan *et al.*, 1994)		AF416464						
AF416472	Son La region		VN	(Bell *et al.*, 2002)		AF416472						
AF416473	Dien Bien Phu region		VN	(Bell *et al.*, 2002)		AF416473						
AF416474	Bac Ninh region		VN	(Bell *et al.*, 2002)		AF416474						

AF416475	Hue region		VN	(Bell *et al.*, 2002)		AF416475					
AF416476	Buon Ma Thout Region		VN	(Bell *et al.*, 2002)		AF416476					
AF416477	Da Nang region		VN	(Bell *et al.*, 2002)		AF416477					
AF416478	Ho Chi Minh City region		VN	(Bell *et al.*, 2002)		AF416478					
AF416479	Yen Bai region		VN	(Bell *et al.*, 2002)		AF416479					
Q279	**Q279**	**C1**	**WS**	**TS**	**1989**	**KM607672**	**KM607801**	**KM607528**	**KM607235**	**KM607093**	**KM607380**
Q280	Q280		WS	TS	1989						KM607381
Q281	**Q281**	**C1**	**WS**	**TS**	**1989**	**KM607673**	**KM607802**	**KM607529**	**KM607236**	**KM607094**	**KM607382**
ABTV1	**Q767**		**MY**	**(Sharman *et al.*, 2008b)**		**EF546813**	**EF546809**	**EF546810**	**EF546811**	**EF546812**	**EF546808**
ABTV2	**Q1108**		**PH**	**(Sharman *et al.*, 2008b)**		**EF546807**	**EF546803**	**EF546804**	**EF546805**	**EF546806**	**EF546802**
ABTV3	Malaysian strain		MY	(Su *et al.*, 2003)					AF102148		

Supplementary Table 3.2: Summary of reassortment detected in BBTV. All detection methods are shown with their corresponding P-values with the most significant shown in bold. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence. Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # Trace evidence was identified for this sequence.

Event #	Reassorted component	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
1	U3	627-TW-1996-D8	24tw-TW 48pk-PK-2007 49pk-PK-2007 60pk-PK 64in-IN-2009 9-150510-EG-2010 B2820-AU-2011 B2823-AU-2011 B2827-AU-2011 B2834-AU-2011 TOS19-TO-2010 TOS85-TO-2010 TOS88-TO-2010 19rw-RW-2009-C2 20rw-RW-2009-C2 26pk-PK-2004-C2 3in-IN-2007-C2 33in-IN-2002-C2 51in-IN-C2 526-BI-1992-C2 549-BI-1995-C2 65lk-LK-2010-C2 BU1-CD-2012-C2 BU10-CD-2012-C2 BU11-CD-2012-C2 BU17-CD-2012-C2 BU20-CD-2012-C2 BU9-CD-2012-C2 Q524-1-IN-C2 Q524-3-IN-C2 All C1 except 5 *35to-TO-2010-C1* *36to-TO-2010-C1* *38to-TO-2010-C* *TOS56-TO-2010-C1* *TOS83-TO-2010-C1*	625I-TW-1995 625-TW-1996 523-6B-IN-1991-D8 All D2 1/1 All D4 1/1 All D5 16/16 All D6 1/1 All D7 1/1	RGMCS**T**	**$3.51x10^{-135}$**
2	S	Q623-TW-1996-D7	547-BI-1995 602-AU-1996 9-150510-EG-2010 B2820-AU-2011 B2828-AU-2011 B2830-AU-2011 B2834-AU-2011 TOS14-TO-2010 TOS19-TO-2010 TOS43-TO-2010 TOS55-TO-2010 TOS58-TO-2010 TOS59-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS68-TO-2010 TOS70-TO-2010 TOS71-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS88-TO-2010 TOS89-TO-2010 All B1 1/1 All C1 40/40 All C2 33/33 All C3 except 1 *B2819-AU-2011-C3*	All D2 1/1 All D4 1/1 All D6 1/1 All D5 except *25tw-TW-D5*	GMCS**T**	**$8.44x10^{-96}$**
3	N	8-150510-EG-2010-A1	1pk-PK-2004 1429B-AU B2818-AU-2011 B2823-AU-2011 B2833-AU-2011	625-TW-1996 768-PH-1995 Q568-3-ID-1995 All D2 1/1 All D3 1/1	R**G**MCST	**$6.07x10^{-70}$**

			B2844-AU-2011 B2846-AU-2011 KP8-AU-1989 TOS4-TO-2010 TOS40-TO-2010 TOS49-TO-2010 TOS57-TO-2010 TOS72-TO-2010 All B1 1/1 All C1 40/40 All C2 33/33 All C3 22/22	All D4 1/1 All D5 16/16 All D6 1/1 All D7 1/1 All D8 2/2 All E1 1/1		
4	U3	Q529-6-CN-1990	9-150510-EG-2010 TOS19-TO-2010 TOS40-TO-2010 TOS49-TO-2010 TOS72-TO-2010 TOS85-TO-2010 TOS88-TO-2010 B2818-AU-2011 B2820-AU-2011 B2833-AU-2011 B2834-AU-2011 34eg-EG-1997 Q524-3-IN-C2 3in-IN-2007-C2 33in-IN-2002-C2 51in-IN-C2 65lk-LK-2010-C2 BU10-CD-2012-C2 BU11-CD-2012-C2 BU17-CD-2012-C2 Q524-1-IN-C2 All C3 22/22 All C1 except 5 *35to-TO-2010-C1* *36to-TO-2010-C1* *38to-TO-2010-C1* *TOS56-TO-2010-C1* *TOS83-TO-2010-C1*	Q529-4-CN-1990-E1	RGMCS**T**	**1.17x10^{-86}**
5	C	24tw-TW	1pk-PK-2004 1429B-AU 9-150510-EG-2010 B2818-AU-2011 B2820-AU-2011 B2828-AU-2011 B2832-AU-2011 B2833-AU-2011 B2846-AU-2011 KP7-AU-1989 KP8-AU-1989 TOS15-TO-2010 TOS4-TO-2010 TOS40-TO-2010 TOS49-TO-2010 TOS57-TO-2010 TOS62-TO-2010 TOS72-TO-2010 TOS76-TO-2010 All B1 1/1 All C1 40/40 All C2 33/33 All C3 22/22	25tw-TW-D5 765tw-TW-1996-D5 MS7-PH-2008-D5 Q1160-TW-1995-D5 Q624-TW-1996-D5 All A1 1/1 All D6 1/1 All D7 1/1	RGMCS**T**	**1.49x10^{-69}**
6	M	TOS88-TO-2010	1pk-PK-2004 22in-IN 547-BI-1995 527-US-1992-C1 KP9-US-1990-C1 Q279-WS-1989-C1 All C2 except 1 *736-4-IN-1997-C2*	TOS19-TO-2010 TOS5-TO-2010 TOS55-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS67-TO-2010 TOS68-TO-2010 TOS71-TO-2010 TOS72-TO-2010 TOS78-TO-2010 TOS80-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS89-TO-2010 37to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 41to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1	RGMCS**T**	**3.38x10^{-25}**

				Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1		
7	N	1429A-AU	602-AU-1996 B2823-AU-2011 B2830-AU-2011 1900A-AU-2006-C3 B2817-AU-2011-C3 B2819-AU-2011-C3 B2822-AU-2011-C3 KP14-AU-2009-C3 KP6-AU-2011-C3	TOS14-TO-2010 TOS55-TO-2010 TOS58-TO-2010 TOS61-TO-2010 TOS68-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS87-TO-2010 All C1 except 9 *44to-TO-2010-C1* *45to-TO-2010-C1* *46to-TO-2010-C* *Q276-TO-1989-C1* *Q277-TO-1989-C1* *Q279-WS-1989-C1* *TOS60-TO-2010-C1* *TOS65-TO-2010-C1* *TOS83-TO-2010-C1*	RGMCS**T**	**1.55x10^{-25}**
8	S	B2819-AU-2011-C3	Unknown	602-AU-1996 737-AU-1997 B2830-AU-2011 B2834-AU-2011 TOS19-TO-2010 TOS43-TO-2010 TOS55-TO-2010 TOS58-TO-2010 TOS59-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS68-TO-2010 TOS70-TO-2010 TOS71-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 36to-TO-2010-C1 37to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 41to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1 736-4-IN-1997-C2 All C3 except 6 *B2817-AU-2011-C3* *B2819-AU-2011-C3* *B2824-AU-2011-C3* *B2829-AU-2011-C3*	R**G**MCST	**2.83x10^{-24}**

				B2845-AU-2011-C3 *KP15-AU-2009-C3*		
9	U3	523-6B-IN-1991-D8	22in-IN 34eg-EG-1997 6us-US 602-AU-1996 64in-IN-2009 9-150510-EG-2010 B2818-AU-2011 B2820-AU-2011 B2823-AU-2011 B2827-AU-2011 B2830-AU-2011 B2833-AU-2011 B2834-AU-2011 Q524-2-IN TOS14-TO-2010 TOS19-TO-2010 TOS4-TO-2010 TOS40-TO-2010 TOS43-TO-2010 TOS49-TO-2010 TOS55-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS67-TO-2010 TOS68-TO-2010 TOS72-TO-2010 TOS78-TO-2010 TOS80-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS88-TO-2010 TOS89-TO-2010 All B1 1/1 All C1 40/40 19rw-RW-2009-C2 20rw-RW-2009-C2 3in-IN-2007-C2 33in-IN-2002-C2 51in-IN-C2 65lk-LK-2010-C2 BU1-CD-2012-C2 BU10-CD-2012-C2 BU11-CD-2012-C2 BU12-CD-2012-C2 BU17-CD-2012-C2 BU20-CD-2012-C2 BU6-CD-2012-C2 BU9-CD-2012-C2 Q524-1-IN-C2 Q524-3-IN-C2 All C3 22/22	625-TW-1996 MP1-TW-1996 MS14-PH-2008 Q568-3-ID-1995 All D2 1/1 All D4 1/1 All D5 16/16 All D6 1/1 All D7 1/1	RGMCS**T**	**$7.04x10^{-78}$**
10	U3	24tw-TW	34eg-EG-1997 9-150510-EG-2010 B2818-AU-2011 B2820-AU-2011 B2833-AU-2011 B2834-AU-2011 TOS19-TO-2010 TOS72-TO-2010 TOS85-TO-2010 All B1 1/1 37to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 41to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1	Q568-3-ID-1995 All D2 1/1 All D4 1/1 All D5 16/16 All D6 1/1 All D7 1/1	RGMCS**T**	**$2.46x10^{-81}$**

			TOS91-TO-2010-C1 TOS93-TO-2010-C1 19rw-RW-2009-C2 20rw-RW-2009-C2 3in-IN-2007-C2 33in-IN-2002-C2 51in-IN-C2 65lk-LK-2010-C2 BU1-CD-2012-C2 BU10-CD-2012-C2 BU11-CD-2012-C2 BU12-CD-2012-C2 BU13-CD-2012-C2 BU17-CD-2012-C2 BU20-CD-2012-C2 BU6-CD-2012-C2 BU7-CD-2012-C2 BU9-CD-2012-C2 Q524-1-IN-C2 Q524-3-IN-C2 All C3 22/22			
11	N	625I-TW-1995	TOS19-TO-2010 TOS5-TO-2010 TOS85-TO-2010 TOS88-TO-2010 All C1 40/40	768-PH-1995 Q529-1-CN-1990 625-TW-1996 All D1 1/1 All D2 1/1 All D3 1/1 All D4 1/1 All D6 1/1 All D8 2/2 All E1 1/1 All D5 except 3 *522B-PH-1991-D5* *571-1-PH-1993-D5* *571-2-PH-1993-D5*	RGMCS**T**	**1.53x10^{-55}**
12	M	736-4-IN-1997-C2	TOS15-TO-2010 TOS19-TO-2010 TOS4-TO-2010 TOS43-TO-2010 TOS5-TO-2010 TOS55-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS68-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 36to-TO-2010-C1 37to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1	547-BI-1995 All C2 except 2 *3in-IN-2007-C2* *736-4-IN-1997-C2*	RGMCS**T**	**9.34x10^{-23}**
13	M	B2833-AU-2011	1429B-AU KP8-AU-1989 BU14-CD-2012-C2	TOS4-TO-2010 TOS58-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS87-TO-2010 TOS22-TO-2010-C1 35to-TO-2010-C1 37to-TO-2010-C1 38to-TO-2010-C1	RM**S**	**3.33x10^{-22}**

				39to-TO-2010-C1 40to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1		
14	R	62cn-CN-D2	21cn-CN-D1	768-PH-1995 Q568-3-ID-1995 625-TW-1996 All D4 1/1 All D6 1/1 All D7 1/1 All D8 2/2 All D5 except 3 *522B-PH-1991-D5* *571-1-PH-1993-D5* *571-2-PH-1993-D5*	RGMC**S**T	**1.59x10^{-15}**
15	N	62cn-CN-D2	21cn-CN-D1	768-PH-1995 625-TW-1996 All D3 1/1 All D4 1/1 All D8 2/2 All D5 except 3 *522B-PH-1991-D5* *571-1-PH-1993-D5* *571-2-PH-1993-D5*	GMS**T**	**3.85x10^{-27}**
16	U3	9-150510-EG-2010 527-US-1992-C1$^{\#}$ KP9-US-1990-C1$^{\#}$ Q279-WS-1989-C1 Q281-WS-1989-C1	TOS19-TO-2010 TOS4-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS85-TO-2010 All C1 except 8 *43to-TO-2010-C1* *44to-TO-2010-C1* *46to-TO-2010-C1* *527-US-1992-C1* *KP9-US-1990-C1* *Q279-WS-1989-C1* *Q281-WS-1989-C1* *TOS83-TO-2010-C1*	All C2 except 8 *26pk-PK-2004-C2* *3in-IN-2007-C2* *33in-IN-2002-C2* *51in-IN-C2* *65lk-LK-2010-C2* *736-4-IN-1997-C2* *KP5-LK-2003-C2* *Q524-1-IN-C2*	RGM**S**T	**7.39x10^{-14}**
17	S	8-150510-EG-2010-A1	1pk-PK-2004 9-150510-EG-2010 TOS15-TO-2010 TOS19-TO-2010 TOS5-TO-2010 TOS62-TO-2010 TOS85-TO-2010 All B1 1/1 All C1 40/40 All C2 except 1 *736-4-IN-1997-C2*	5tw-TW 768-PH-1995 All D4 1/1 All D5 16/16 All D6 1/1 All D8 2/2	RGBMS**T**	**1.34x10^{-86}**
18	M	TOS14-TO-2010	1pk-PK-2004 1429B-AU 22in-IN 547-BI-1995 KP9-US-1990-C1 Q279-WS-1989-C1 All C2 except 1 *736-4-IN-1997-C2*	TOS19-TO-2010 TOS4-TO-2010 TOS55-TO-2010 TOS58-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS67-TO-2010 TOS68-TO-2010 TOS72-TO-2010 TOS78-TO-2010 TOS80-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 36to-TO-2010-C1 37to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1	RGBMCS**T**	**3.37x10^{-19}**

				41to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1		
19	S	TOS71-TO-2010	602-AU-1996 B2820-AU-2011 B2823-AU-2011 B2828-AU-2011 B2830-AU-2011 B2834-AU-2011 19rw-RW-2009-C2 20rw-RW-2009-C2 3in-IN-2007-C2 47mw-MW-2008-C2 526-BI-1992-C2 549-BI-1995-C2 550-CG-1995-C2 BU1-CD-2012-C2 BU11-CD-2012-C2 BU12-CD-2012-C2 BU13-CD-2012-C2 BU14-CD-2012-C2 BU15-CD-2012-C2 BU16-CD-2012-C2 BU17-CD-2012-C2 BU18-CD-2012-C2 BU19-CD-2012-C2 BU2-CD-2012-C2 BU20-CD-2012-C2 BU6-CD-2012-C2 BU7-CD-2012-C2 Q553-LK-1995-C2 All C3 except 1 *B2819-AU-2011-C3*	TOS14-TO-2010 TOS19-TO-2010 TOS55-TO-2010 TOS61-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS68-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 All C1 except 8 *43to-TO-2010-C1* *44to-TO-2010-C1* *527-US-1992-C1* *KP9-US-1990-C1* *Q279-WS-1989-C1* *Q281-WS-1989-C1* *TOS90-TO-2010-C1* *TOS93-TO-2010-C1*	RGB**T**	**1.10×10^{-11}**
20	U3	44to-TO-2010-C1 45to-TO-2010-C1 46to-TO-2010-C1	B2818-AU-2011 B2823-AU-2011 B2830-AU-2011 B2833-AU-2011 B2834-AU-2011 All C3 except 2 *1900B-AU-2006-C3* *KP17-AU-2010-C3*	TOS93-TO-2010-C1	RGBMS**T**	**2.76×10^{-22}**
21	C	KP9-US-1990-C1 527-US-1992-C1 Q279-WS-1989-C1 Q281-WS-1989-C1	TOS19-TO-2010 TOS4-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS67-TO-2010 TOS68-TO-2010 TOS80-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 36to-TO-2010-C1 37to-TO-2010-C1 41to-TO-2010-C1 KP4-TO-1990-C1 Q570-TO-1990-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS93-TO-2010-C1	22in-IN 33in-IN-2002-C2 47mw-MW-2008-C2 51in-IN-C2 526-BI-1992-C2 548-BI-1995-C2 550-CG-1995-C2 65lk-LK-2010-C2 BU1-CD-2012-C2 BU10-CD-2012-C2 BU12-CD-2012-C2 BU13-CD-2012-C2 BU15-CD-2012-C2 BU16-CD-2012-C2 BU17-CD-2012-C2 BU18-CD-2012-C2 BU19-CD-2012-C2 BU6-CD-2012-C2 BU7-CD-2012-C2 KP5-LK-2003-C2 Q553-LK-1995-C2 20rw-RW-2009-C2 19rw-RW-2009-C2	RBMS**T**	**8.82×10^{-35}**

22	U3	Q529-4-CN-1990-E1 Q529-2-CN-1990	TOS14-TO-2010 TOS19-TO-2010 TOS88-TO-2010 TOS91-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 42to-TO-2010-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS39-TO-2010-C1 TOS48-TO-2010-C1 TOS60-TO-2010-C1 TOS65-TO-2010-C1 TOS90-TO-2010-C1 TOS93-TO-2010-C1 33in-IN-2002-C2 51in-IN-C2 1900A-AU-2006-C3 1900B-AU-2006-C3 KP17-AU-2010-C3 KP18-AU-2010-C3	625I-TW-1995 625-TW-1996 All D1 1/1 All D2 1/1 All D4 1/1 All D6 1/1 All D5 except 5 *522A-PH-1991-D5* *571-2-PH-1993-D5* *MS17-PH-2008-D5* *MS6-PH-2008-D5* *MS7-PH-2008-D5*	**M**CS	**4.78x10^{-05}**
23	U3	B2818-AU-2011 B2823-AU-2011 B2827-AU-2011 B2833-AU-2011 B2834-AU-2011 All C3 22/22	TOS19-TO-2010 TOS40-TO-2010 TOS62-TO-2010 TOS72-TO-2010 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 42to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS39-TO-2010-C1 TOS42-TO-2010-C1 TOS48-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS91-TO-2010-C1	27-IN-2006 64in-IN-2009 526-BI-1992-C2 548-BI-1995-C2 549-BI-1995-C2 736-4-IN-1997-C2 BU10-CD-2012-C2 BU11-CD-2012-C2 BU14-CD-2012-C2 BU16-CD-2012-C2 BU17-CD-2012-C2 BU2-CD-2012-C2 BU7-CD-2012-C2 BU9-CD-2012-C2 Q524-1-IN-C2 Q524-3-IN-C2 Q553-LK-1995-C2	RGBMC**S**	**3.03x10^{-14}**
24	N	1429B-AU KP8-AU-1989	602-AU-1996 737-AU-1997 B2818-AU-2011 B2826-AU-2011 B2830-AU-2011 B2834-AU-2011 B2846-AU-2011 All C3 22/22	Unknown	RGBMC**T**	**1.77x10^{-17}**
25	S	523-6A-IN-1991	29mm-MM 30mm-MM 31mm-MM 64in-IN-2009 9-150510-EG-2010 B2820-AU-2011 B2826-AU-2011 B2828-AU-2011 B2832-AU-2011 B2834-AU-2011 B2846-AU-2011 TOS15-TO-2010 TOS19-TO-2010 TOS5-TO-2010 TOS62-TO-2010 TOS85-TO-2010 All B1 1/1 All C1 40/40 All C2 33/33 All C3 except 1 *B2819-AU-2011-C3*	10ph-PH 13ph-PH 14ph-PH 15ph-PH 16id-ID 17id-ID 18id-ID 5tw-TW 7jp-JP 768-PH-1995 8jp-JP 9jp-JP All D1 1/1 All D4 1/1 All D5 16/16 All D6 1/1 All D8 2/2	RGBMCS**T**	**2.15x10^{-64}**
26	M	B2834-AU-2011	TOS15-TO-2010 TOS19-TO-2010 TOS4-TO-2010 TOS5-TO-2010 TOS55-TO-2010 TOS57-TO-2010	1pk-PK-2004 9-150510-EG-2010 Q281-WS-1989-C1 BU17-CD-2012-C2 All C2 except 2 *26pk-PK-2004-C2*	RGBMCS**T**	**8.43x10^{-20}**

			TOS58-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS68-TO-2010 TOS72-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 36to-TO-2010-C1 37to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 41to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1	*736-4-IN-1997-C2*		
27	R	KP8-AU-1989 1429B-AU KP7-AU-1989	B2818-AU-2011 B2820-AU-2011 B2826-AU-2011 B2828-AU-2011 B2832-AU-2011 B2834-AU-2011 B2846-AU-2011 All C3 22/22	Unknown	RBMCS**T**	**$2.13x10^{-21}$**
28	M	TOS40-TO-2010 TOS49-TO-2010 42to-TO-2010-C1 43to-TO-2010-C1 44to-TO-2010-C1 45to-TO-2010-C1 46to-TO-2010-C1 TOS39-TO-2010-C1 TOS48-TO-2010-C1 TOS90-TO-2010-C1	547-BI-1995 All C2 except 3 *26pk-PK-2004-C2* *736-4-IN-1997-C2* *BU17-CD-2012-C2*	TOS15-TO-2010 TOS19-TO-2010 TOS43-TO-2010 TOS55-TO-2010 TOS62-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS68-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS89-TO-2010 37to-TO-2010-C1 35to-TO-2010-C1 36to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1	RGBMCS**T**	**$6.63x10^{-18}$**
29	N	527-US-1992-C1	TOS19-TO-2010	1pk-PK-2004	GBMCS**T**	**$8.44x10^{-09}$**

		KP9-US-1990-C1 Q279-WS-1989-C1 Q281-WS-1989-C1	TOS4-TO-2010 TOS5-TO-2010 TOS55-TO-2010 TOS57-TO-2010 TOS63B-TO-2010 TOS64-TO-2010 TOS67-TO-2010 TOS68-TO-2010 TOS71-TO-2010 TOS72-TO-2010 TOS80-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS87-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS25-TO-2010-C1 TOS29-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1	22in-IN 547-BI-1995 All C2 except 1 *736-4-IN-1997-C2*		
30	N	63cn-CN-D3	23cn-CN-2008-D4	Unknown	GB**S**	**5.89x10^{-05}**
31	S	24tw-TW	1pk-PK-2004 32fj-FJ 64in-IN-2009 9-150510-EG-2010 B2820-AU-2011 B2826-AU-2011 B2828-AU-2011 B2832-AU-2011 B2846-AU-2011 TOS15-TO-2010 TOS19-TO-2010 TOS34-TO-2010 TOS45to-TO-2010 TOS46to-TO-2010 TOS5-TO-2010 TOS62-TO-2010 TOS85-TO-2010 TOS88-TO-2010 All B1 1/1 All C1 except 8 *42to-TO-2010-C1* *43to-TO-2010-C1* *44to-TO-2010-C1* *45to-TO-2010-C1* *46to-TO-2010-C1* *TOS39-TO-2010-C1* *TOS48-TO-2010-C1* *TOS90-TO-2010-C1* All C2 except 1 *33in-IN-2002-C2* All C3 except 1 *B2819-AU-2011-C3*	14ph-PH 15ph-PH 16id-ID 17id-ID 18id-ID 5tw-TW 625-TW-1996 768-PH-1995 All D4 1/1 All D5 16/16 All D6 1/1 All D8 2/2 All E1 1/1	RGBMS**T**	**8.39x10^{-66}**
32	U3	63cn-CN-D3	21cn-CN-D1	MS14-PH-2008 All D2 1/1 All D4 1/1 All D5 except 4 *522A-PH-1991-D5* *522B-PH-1991-D5* *765tw-TW-1996-D5* *Q624-TW-1996-D5*	RGBMCS**T**	**2.47x10^{-21}**
33	M	B2818-AU-2011[#] B2820-AU-2011 B2828-AU-2011[#] B2846-AU-2011 All C3 22/22	TOS19-TO-2010 TOS5-TO-2010 TOS55-TO-2010 TOS57-TO-2010 TOS58-TO-2010 TOS64-TO-2010	547-BI-1995 All C2 except 2 *26pk-PK-2004-C2* *736-4-IN-1997-C2*	RGBMCS**T**	**6.87x10^{-20}**

			TOS68-TO-2010 TOS72-TO-2010 TOS78-TO-2010 TOS82-TO-2010 TOS85-TO-2010 TOS89-TO-2010 35to-TO-2010-C1 36to-TO-2010-C1 37to-TO-2010-C1 38to-TO-2010-C1 39to-TO-2010-C1 40to-TO-2010-C1 41to-TO-2010-C1 536-TO-1993-C1 KP4-TO-1990-C1 Q276-TO-1989-C1 Q277-TO-1989-C1 Q278-TO-1989-C1 Q570-TO-1990-C1 TOS16-TO-2010-C1 TOS2-TO-2010-C1 TOS20-TO-2010-C1 TOS21-TO-2010-C1 TOS22-TO-2010-C1 TOS42-TO-2010-C1 TOS56-TO-2010-C1 TOS60-TO-2010-C1 TOS63A-TO-2010-C1 TOS65-TO-2010-C1 TOS83-TO-2010-C1 TOS91-TO-2010-C1 TOS93-TO-2010-C1			
34	S	1429A-AU	TOS63A-TO-2010-C1	33in-IN-2002-C2	RM**T**	**8.15x10^{-04}**
35	U3	3in-IN-2007-C2	Unknown	65lk-LK-2010-C2	GB**S**	**3.48x10^{-16}**
36	M	602-AU-1996 737-AU-1997 B2826-AU-2011[T] B2830-AU-2011 B2832-AU-2011	TOS22-TO-2010-C1	3in-IN-2007-C2	RGS**T**	**3.84x10^{-12}**
37	C	MP2-TW-1996-D6	520-ID-1995-D5 Q568-1-ID-1995-D5 Q568-3-ID-1995	Unknown	R**G**BS	**3.95x10^{-06}**
38	C	BU11-CD-2012-C2	TOS4-TO-2010 TOS58-TO-2010 TOS62-TO-2010 TOS71-TO-2010 TOS83-TO-2010-C1 TOS93-TO-2010-C1 35to-TO-2010-C1 Q278-TO-1989-C1	Q553-LK-1995-C2	RB**M**S	**9.33x10^{-03}**
39	C	9-150510-EG-2010 KP5-LK-2003	Unknown	548-BI-1995	M**S**	**1.51X10^{-08}**
40	U3	8-150510-EG-2010 A1	571-1-PH-1993-D5	Unknown	RGMCS**T**	**9.02x10^{-8}**

Supplementary Table 3.3: Summary of recombination events detected in DNA-R components. All detection methods are shown with their corresponding P-values, with the most significant shown in bold. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence. Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # Trace evidence was identified for this sequence.

Event #	Breakpoints in Alignment	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
R1	325-732	AF416476-R-VN AF416477-R-VN[#] AF416478-R-VN	Unknown	10ph-R-PH 11vn-R-VN 13ph-R-PH 14ph-R-PH 15ph-R-PH 16id-R-ID 17id-R-ID 18id-R-ID 24tw-R-TW 523-6A-R-IN-1991 7jp-R-JP 768-R-PH-1995 8jp-R-JP 9jp-R-JP AB108454-R-JP AB108455-R-JP AB108457-R-JP AB108458-R-JP AF110266-R-CN AF246123-R-CN AF416468-R-TW AF416469-R-PH JN003631-R-ID-2010 JN003632-R-ID-2010 JN003633-R-ID-2010 Q568-3-R-ID-1995 627-R-TW-1996-D8 All D5 except 6 *571-2-R-PH-1993-D5* *MS15-R-PH-2008-D5* *MS16-R-PH-2008-D5* *MS18-R-PH-2008-D5* *Q1160-R-TW-1995-D5* *Q624-R-TW-1996-D5)*	**M**CT	**1.70x10^{-05}**
R2	1102-46	MP2-R-TW-1996-D6	Unknown	8-150510-R-EG-2010-A1 10ph-R-PH 13ph-R-PH 14ph-R-PH 15ph-R-PH 16id-R-ID 17id-R-ID 18id-R-ID 24tw-R-TW 523-6A-R-IN-1991 625-R-TW-1996 7jp-R-JP 768-R-PH-1995 8jp-R-JP 9jp-R-JP AB108454-R-JP AB108455-R-JP AB108457-R-JP AB108458-R-JP AF416468-R-TW AF416469-R-PH AF416476-R-VN JN003631-R-ID-2010 JN003632-R-ID-2010 JN003633-R-ID-2010 Q568-3-R-ID-1995 All D5 16/16 All D7 1/1 All D8 2/2	**R**GBT	**8.04x10^{-05}**
R3	1088-378	5tw-R-TW 625I-R-TW-1995[#]	Q279-R-WS-1989-C1 Q281-R-WS-1989-C1	21cn-R-CN-D1 62cn-R-CN-D2 63cn-R-CN-D3	**M**CS	**3.68x10^{-04}**
R4	169-312	21cn-R-CN-D1 62cn-N-R-CN-D2 63cn-R-CN-D3	Q281-R-WS-1989-C1 TOS93-R-TO-2010-C1	Unknown	**R**GB	**2.95x10^{-03}**
R6	97-579	6us-R-US[#] 527-R-US-1992-C1[#] KP9-R-US-1990-C1[#] Q279-R-WS-1989-C1[#] Q281-R-WS-1989-C1	TOS91-R-TO-2010-C1	33in-R-IN-2002-C2	MC**S**	**3.70x10^{-03}**

Supplementary Table 3.4: Summary of recombination events detected in DNA-U3 components. All detection methods are shown with their corresponding P-values, with the most significant shown in bold. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence. Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # Trace evidence was identified for this sequence.

Event #	Breakpoints in Alignment	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
U2	1558-1778	27in-U3-IN-2006 48pk-U3-PK-2007# 49pk-U3-PK-2007# 547-U3-BI-1995# 60pk-U3-PK 64in-U3-IN-2009# AY884172-U3-IN# AY996563-U3-PK-2007# TOS4-U3-TO-2010 TOS58-U3-TO-2010 TOS59-U3-TO-2010 TOS61-U3-TO-2010 TOS62-U3-TO-2010 TOS69-U3-TO-2010 TOS71-U3-TO-2010 TOS74-U3-TO-2010 35to-U3-TO-2010-C1 36to-U3-TO-2010-C1 38to-U3-TO-2010-C1 TOS56-U3-TO-2010-C1 TOS83-U3-TO-2010-C1 523-6B-U3-IN-1991-D8# All C2 except 7 *3in-U3-IN-2007-C2* *33in-U3-IN-2002-C2* *51in-U3-IN-C2* *65lk-U3-LK-2010-C2* *Q524-1-U3-IN-C2* *Q524-3-U3-IN-C2*	Unknown	22in-U3-IN 24tw-U3-TW 28in-U3-IN-2012 34eg-U3-EG-1997 6us-U3-US 602-U3-AU-1996 9-150510-U3-EG-2010 AY884173-U3-IN B2818-U3-AU-2011 B2823-U3-AU-2011 B2827-U3-AU-2011 B2828-U3-AU-2011 B2830-U3-AU-2011 B2833-U3-AU-2011 B2834-U3-AU-2011 EU046323-U3-IN Q524-2-U3-IN Q529-6-U3-CN-1990 TOS14-U3-TO-2010 TOS19-U3-TO-2010 TOS40-U3-TO-2010 TOS43-U3-TO-2010 TOS49-U3-TO-2010 TOS55-U3-TO-2010 TOS63B-U3-TO-2010 TOS64-U3-TO-2010 TOS67-U3-TO-2010 TOS68-U3-TO-2010 TOS72-U3-TO-2010 TOS78-U3-TO-2010 TOS79-U3-TO-2010 TOS80-U3-TO-2010 TOS82-U3-TO-2010 TOS85-U3-TO-2010 TOS87-U3-TO-2010 TOS88-U3-TO-2010 TOS89-U3-TO-2010 All A1 1/1 3in-U3-IN-2007-C2 33in-U3-IN-2002-C2 51in-U3-IN-C2 65lk-U3-LK-2010-C2 Q524-1-U3-IN-C2 Q524-3-U3-IN-C2 All C3 22/22 627-U3-TW-1996-D8 All C1 except 6 *35to-U3-TO-2010-C1* *36to-U3-TO-2010-C1* *38to-U3-TO-2010-C1* *44to-U3-TO-2010-C1* *TOS56-U3-TO-2010-C1* *TOS83-U3-TO-2010-C1*	R**G**MCST	**1.17x10^{-17}**
U4	1674-17	25tw-U3-TW-D5 MS14-U3-PH-2008	Unknown	625-U3-TW-1996 AY996563-U3-PK-2007 DQ826392-U3-TW GU559703-U3-CN-2008 Q568-3-U3-ID-1995 BU10-U3-CD-2012-C2 BU17-U3-CD-2012-C2 All D2 1/1 All D4 1/1 All D7 1/1 All D5 except 2 *571-2-U3-PH-1993-D5* *25-TW-U3-TW-D5 (recombinant)*	R**G**MST	7.77x10^{-12}
U5	1706-214	AY996563-U3-PK-2007 19rw-U3-RW-2009-C2# 20rw-U3-RW-2009-C2# BU1-U3-CD-2012-C2# BU10-U3-CD-2012-C2	625-U3-TW-1996 All D2 1/1 522A-U3-PH-1991-D5 522B-U3-PH-1991-D5 Q1160-U3-TW-1995-D5	24tw-U3-TW 34eg-U3-EG-1997 6us-U3-US 602-U3-AU-1996 9-150510-U3-EG-2010	R**G**T	4.07x10^{-08}

		BU17-U3-CD-2012-C2 BU9-U3-CD-2012-C2[#]	Q568-1-U3-ID-1995-D5	AY884173-U3-IN B2818-U3-AU-2011 B2820-U3-AU-2011 B2823-U3-AU-2011 B2827-U3-AU-2011 B2828-U3-AU-2011 B2830-U3-AU-2011 EU046323-U3-IN Q524-2-U3-IN Q529-2-U3-CN-1990 Q529-6-U3-CN-1990 TOS14-U3-TO-2010 TOS40-U3-TO-2010 TOS43-U3-TO-2010 TOS49-U3-TO-2010 TOS55-U3-TO-2010 TOS63B-U3-TO-2010 TOS64-U3-TO-2010 TOS67-U3-TO-2010 TOS68-U3-TO-2010 TOS72-U3-TO-2010 TOS78-U3-TO-2010 TOS79-U3-TO-2010 TOS80-U3-TO-2010 TOS82-U3-TO-2010 TOS85-U3-TO-2010 TOS87-U3-TO-2010 TOS89-U3-TO-2010 All A1 1/1 3in-U3-IN-2007-C2 33in-U3-IN-2002-C2 51in-U3-IN-C2 65lk-U3-LK-2010-C2 Q524-1-U3-IN-C2 Q524-3-U3-IN-C2 627-U3-TW-1996-D8 All E1 1/1 All C1 except 9 *35to-U3-TO-2010-C1* *36to-U3-TO-2010-C1* *38to-U3-TO-2010-C1* *TOS25-U3-TO-2010-C1* *TOS29-U3-TO-2010-C1* *TOS42-U3-TO-2010-C1* *TOS56-U3-TO-2010-C1* *TOS83-U3-TO-2010-C1* *TOS90-U3-TO-2010-C1* All C3 except 2 *482-98-U3-AU-1998-C3* *KP18-U3-AU-2010-C3*		
U6	1321-1789	TOS93-U3-TO-2010-C1	Unknown	Q529-6-U3-CN-1990 TOS55-U3-TO-2010 TOS80-U3-TO-2010 TOS65-U3-TO-2010-C1	B**M**T	**1.22x10^{-04}**
U7	1585-353	5tw-U3-TW	Unknown	625-U3-TW-1996 625I-U3-TW-1995 DQ826392-U3-TW FJ773283-U3-TW GU559702-U3-CN-2009 GU559703-U3-CN-2008 GU559704-U3-CN-2009 GU559705-U3-CN-2009 GU559706-U3-CN-2009 HM212635-U3-CN-2009 MP1-U3-TW-1996 MS14-U3-PH-2008 Q529-2-U3-CN-1990 Q568-3-U3-ID-1995 All D1 1/1 All D2 1/1 All D3 1/1 All D4 1/1 All D5 16/16 All D6 1/1 All D7 1/1 All E1 1/1	GMS**T**	**5.93x10^{-19}**
U8	1608-38	625-U3-TW-1996 625I-U3-TW-1995[#] DQ826392-U3-TW FJ773283-U3-TW[#] GU559703-U3-CN-2008 MP1-U3-TW-1996 MS14-U3-PH-2008 Q568-3-U3-ID-1995 All D2 1/1 All D4 1/1	Unknown	GU559702-U3-CN-2009 GU559704-U3-CN-2009 GU559705-U3-CN-2009 GU559706-U3-CN-2009 HM212635-U3-CN-2009 All D1 1/1 All D3 1/1	**R**GMCST	**7.39x10^{-11}**

		All D5 16/16 All D6 1/1 All D7 1/1				
U10	356*-567	5tw-U3-TW	TOS19-U3-TO-2010 TOS58-U3-TO-2010 TOS64-U3-TO-2010 All A1 1/1 38to-U3-TO-2010-C1 43to-U3-TO-2010-C1 527-U3-US-1992-C1 536-U3-TO-1993-C1 KP4-U3-TO-1990-C1 Q279-U3-WS-1989-C1 Q570-U3-TO-1990-C1 TOS16-U3-TO-2010-C1 TOS39-U3-TO-2010-C1 TOS42-U3-TO-2010-C1 TOS65-U3-TO-2010-C1 TOS90-U3-TO-2010-C1	571-1-U3-PH-1993-D5	**R**GBST	**7.21x10^{-05}**
U12	374-534	Q529-4-U3-CN-1990-E1 Q529-2-U3-CN-1990[#]	TOS64-U3-TO-2010 TOS39-U3-TO-2010-C1 43to-U3-TO-2010-C1	Unknown	R**G**B	**1.86x10^{-03}**
U17	460-1011	AY884173-U3-IN Q529-6-U3-CN-1990 33in-U3-IN-2002-C2 51in-U3-IN-C2 65lk-U3-LK-2010-C2 Q524-1-U3-IN-C2 Q524-3-U3-IN-C2[#]	Unknown	8-150510-U3-EG-2010-A1	**M**CST	**6.68x10^{-05}**
U19	357-563	66in-U3-IN-2012-B1	21cn-U3-CN-D1	482-98-U3-AU-1998-C3	RMC**S**	**1.07x10^{-04}**
U21	1010-1558	8-150510-U3-EG-2010-A1	Unknown	B2818-U3-AU-2011 B2820-U3-AU-2011 B2823-U3-AU-2011 B2827-U3-AU-2011 24-TW-U3-TW 602-U3-AU-1996 B2828-U3-AU-2011 B2830-U3-AU-2011 B2833-U3-AU-2011 B2834-U3-AU-2011 44to-U3-TO-2010-C1 45to-U3-TO-2010-C1 46to-U3-TO-2010-C1 All C3 except 2 *KP14-U3-AU-2009-C3* *KP17-U3-AU-2010-C3*	MC**S**	**1.85x10^{-05}**
U22	371-504	TOS43-U3-TO-2010	Unknown	TOS42-U3-TO-2010-C1	R**G**B	**2.47x10^{-03}**

Supplementary Table 3.5: Summary of recombination events detected in DNA-S components. All detection methods are shown with their corresponding P-values, with the most significant shown in bold. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence.Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # - Trace evidence was identified for this sequence.

Event #	Breakpoints in Alignment	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
S1	1193-68	626-S-TW-1996 626M-S-TW-1995 MP2-S-TW-1996-D6	Unknown	10ph-S-PH 13ph -S-PH 14ph -S-PH 15-ph -S-PH 16id-S-ID 17id -S-ID 18id -S-ID 7jp-S-JP 768-S-PH-1995 8jp-S-JP 9jp-S-JP AF148068-S-PH AF148942-S-TW AF238877-S-CN MP1-S-TW-1996 MS14-S-PH-2008 All D2 1/1 All D3 1/1 All D4 1/1 All D5 16/16 All D8 2/2	RG**T**	**3.96x10^{-09}**
S2	362-730	10ph-S-PH 11vn-S-VN[#] 12vn-S-VN[#] 13ph-S-PH 14ph-S-PH 15ph-S-PH 16id-S-ID 17id-S-ID 18id-S-ID 5tw-S-TW 625-S-TW-1996 625I-S-TW-1995 626-S-TW-1996 626M-S-TW-1995 7jp-S-JP 768-S-PH-1995 8jp-S-JP 9jp-S-JP AF148068-S-PH AF148942-S-TW AF148945-S-VN[#] AF238876-S-CN AF238877-S-CN[#] MP1-S-TW-1996 MS14-S-PH-2008 Q529-1-S-CN-1990 Q529-2-S-CN-1990[#] Q529-5-S-CN-1990[#] All D1 1/1 All D2 1/1 All D3 1/1 All D4 1/1 All D5 16/16 All D6 1/1 All D8 2/2 All E1 1/1	B2846-S-AU-2011 BU18-S-CD-2012-C2 BU9-S-CD-2012-C2	Unknown	RGMCS**T**	**3.72x10^{-06}**
S5	421-478	B2846-S-AU-2011	Unknown	1pk-S-PK-2004 2pk-S-PK-2004 AM418565-S-PK-2004 AM418567-S-PK-2004 GQ249344-S-CM-2008 JF755978-S-CM-2008 JF755980-S-MW-2008 JF755981-S-GA-2008 JF755982-S-GA-2008 JF755984-S-CD-2008 JF755986-S-CD-2008 JF755987-S-CD-2008	RG**B**	**6.90x10^{-04}**
S7	392*-514	JF755981-S-GA-2008 JF755984-S-CD-2008	Unknown	B2819-S-AU-2011-C3	RG**B**	**2.90x10^{-03}**
S8	1193-64	5tw-S-TW[#] 625-S-TW-1996[#] 625I-S-TW-1995	Unknown	571-2-S-PH-1993-D5	R**G**B	**7.27x10^{-03}**

Supplementary Table 3.6: Summary of recombination events detected in DNA-M components. All detection methods are shown with their corresponding P-values, with the most significant shown in bold. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence. Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # Trace evidence was identified for this sequence.

Event #	Breakpoints in Alignment	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
M1	406-759	66in-M-IN-2012-B1	1429A-M-AU 547-M-BI-1995 33in-M-IN-2002-C2 548-M-BI-1995-C2 BU12-M-CD-2012-C2 BU13-M-CD-2012-C2 BU17-M-CD-2012-C2	TOS72-M-TO-2010 TOS15-M-TO-2010 TOS19-M-TO-2010 TOS43-M-TO-2010 TOS55-M-TO-2010 TOS57-M-TO-2010 TOS59-M-TO-2010 TOS63B-M-TO-2010 TOS67-M-TO-2010 TOS68-M-TO-2010 TOS69-M-TO-2010 TOS71-M-TO-2010 TOS76-M-TO-2010 TOS78-M-TO-2010 TOS80-M-TO-2010 TOS82-M-TO-2010 TOS85-M-TO-2010 TOS87-M-TO-2010 TOS89-M-TO-2010 B2828-M-AU-2011 B2830-M-AU-2011 B2833-M-AU-2011 B2834-M-AU-2011 B2846-M-AU-2011 35to-M-TO-2010-C1 37to-M-TO-2010-C1 40to-M-TO-2010-C1 41to-M-TO-2010-C1 536-M-TO-1993-C1 KP4-M-TO-1990-C1 Q276-M-TO-1989-C1 Q277-M-TO-1989-C1 Q570-M-TO-1990-C1 TOS16-M-TO-2010-C1 TOS29-M-TO-2010-C1 TOS56-M-TO-2010-C1 TOS63A-M-TO-2010-C1 TOS65-M-TO-2010-C1 TOS83-M-TO-2010-C1 TOS91-M-TO-2010-C1 736-4-M-IN-1997-C2 1900B-M-AU-2006-C3 2557-M-AU-2010-C3 482-98-M-AU-1998-C3 B2845-M-AU-2011-C3 KP14-M-AU-2009-C3 KP17-M-AU-2010-C3	GBMS	**9.97×10^{-05}**
M5	1285-15	ABTV3-M-MY	1pk-M-PK-2004 1429B-M-AU 22in-M-IN 24tw-M-TW 523-6A-M-IN-1991 768-M-PH-1995 AF349568-M-CN AY953429-M-IN HE864318-M-PK HE864319-M-PK KC581796-M-TH-2012 KP7-M-AU-1989 KP8-M-AU-1989 MS14-M-PH-2008 Q529-6-M-CN-1990 Q568-3-M-ID-1995 TOS40-M-TO-2010 TOS49-M-TO-2010 TOS88-M-TO-2010 42to-M-TO-2010-C1 43to-M-TO-2010-C1 45to-M-TO-2010-C1 527-M-US-1992-C1 KP9-M-US-1990-C1 Q281-M-WS-1989-C1	ABTV2-M-PH	GBS	**2.19×10^{-04}**

	TOS39-M-TO-2010-C1
	TOS48-M-TO-2010-C1
	All D1 1/1
	All D2 1/1
	All D3 1/1
	All D4 1/1
	All D8 2/2
	All E1 1/1
	All D5 except 5
	25tw-M-TW-D5
	765-M-TW-1996-D5
	S7-M-PH-2008-D5
	Q1160-M-TW-1995-D5
	Q624-M-TW-1996-D5
	All C2 except 8
	33in-M-IN-2002-C2
	51in-M-IN-C2
	549-M-BI-1995-C2
	736-4-M-IN-1997-C2
	BU14-M-CD-2012-C2
	BU15-M-CD-2012-C2
	BU16-M-CD-2012-C2
	BU20-M-CD-2012-C2

Supplementary Table 3.7: Summary of recombination events detected in DNA-C components. All detection methods are shown with their corresponding P-values, with the most significant shown in bold Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence. Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # Trace evidence was identified for this sequence.

Event #	Breakpoints in Alignment	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
C1	277-780	3in-C-IN-2007-C2	66in-C-IN-2012-B1	1pk-C-PK-2004 1429A-C-AU 1429B-C-AU 2pk-C-PK-2004 24tw-C-TW 547-C-BI-1995 6us-C-US 602-C-AU-1996 737-C-AU-1997 9-150510-C-EG-2010 B2818-C-AU-2011 B2820-C-AU-2011 B2823-C-AU-2011 B2827-C-AU-2011 B2828-C-AU-2011 B2830-C-AU-2011 B2832-C-AU-2011 B2833-C-AU-2011 B2844-C-AU-2011 B2846-C-AU-2011 KP7-C-AU-1989 KP8-C-AU-1989 Q524-2-C-IN TOS14-C-TO-2010 TOS15-C-TO-2010 TOS4-C-TO-2010 TOS40-C-TO-2010 TOS43-C-TO-2010 TOS49-C-TO-2010 TOS55-C-TO-2010 TOS57-C-TO-2010 TOS58-C-TO-2010 TOS59-C-TO-2010 TOS61-C-TO-2010 TOS62-C-TO-2010 TOS63B-C-TO-2010 TOS64-C-TO-2010 TOS67-C-TO-2010 TOS68-C-TO-2010 TOS69-C-TO-2010 TOS70-C-TO-2010 TOS71-C-TO-2010 TOS72-C-TO-2010 TOS76-C-TO-2010 TOS78-C-TO-2010 TOS82-C-TO-2010 TOS87-C-TO-2010 All C3 22/22 All C1 except 1 *TOS93-C-TO-2010-C1* All C2 except 2 *3in-C-IN-2007-C2* *BU10-C-CD-2012-C2*	RGMC**T**	**1.03x10^{-07}**
C2	1068-28	8-150510-C-EG-2010-A1 625-C-TW-1996 765-C-TW-1996-D5 Q624-C-TW-1996-D5 All D6 1/1 All D7 1/1	Unknown	AY264347-C-CN AY266417-C-CN Q568-3-C-ID-1995 MS17-C-PH-2008-D5 Q1160-C-TW-1995-D5 Q568-1-C-ID-1995-D5 523-6B-C-IN-1991-D8 All D3 1/1	**R**GT	**1.41x10^{-06}**
C3	500-1108	Q529-2-C-CN-1990 Q529-4-C-CN-1990-E1	Unknown	21cn-C-CN-D1	MC**S**	**4.42x10^{-06}**
C4	430-482*	526-C-BI-1992-C2	Unknown	BU18-C-CD-2012-C2	GB**T**	**6.65x10^{-03}**

Supplementary Table 3.8: Summary of recombination events detected in DNA-N components. All detection methods are shown with their corresponding P-values, with the most significant shown in bold. Minor parent, parent contributing to the smaller fraction of the sequence. Major parent, parent contributing the larger fraction of sequence. Only one parent and a recombinant need to be in the alignment for a recombinant to be detectable, the sequence listed as Unknown was used to infer the existence of a missing parental sequence. # Trace evidence was identified for this sequence.

Event #	Breakpoints in Alignment	Recombinant Sequence(s)	Sequence(s) used to infer minor parent(s)	Sequence(s) used to infer major parent(s)	Detection Methods	p-value
N1	300-318	TOS53-N-TO-2010	Unknown	1pk-N-PK-2004 1429A-N-AU 1429B-N-AU 22in-N-IN 547-N-BI-1995 602-N-AU-1996 625I-N-TW-1995 626-N-TW-1996 64in-N-IN-2009 737-N-AU-1997 AF238878-N-CN AF238879-N-CN B2818-N-AU-2011 B2823-N-AU-2011 B2826-N-AU-2011 B2830-N-AU-2011 B2833-N-AU-2011 B2834-N-AU-2011 B2844-N-AU-2011 B2846-N-AU-2011 EF470243-N-CN HE864320-N-PK KP8-N-AU-1989 Q529-1-N-CN-1990 Q529-2-N-CN-1990 All A1 1/1 All C1 40/40 All C2 33/33 All D1 1/1 All D2 1/1 All D3 1/1 All D4 1/1 All D6 1/1 All D8 2/2 All C3 except 1 *B2819-N-AU-2011-C3* All ungrouped TO except 1 *TOS53-N-TO-2010 (recombinant)* All D5 except 5 *522B-N-PH-1991-D5* *571-2-N-PH-1993-D5* *MS7-N-PH-2008-D5* *Q568-1-N-ID-1995-D5* *Q624-N-TW-1996-D5*	RG**B**	**2.95x10^{-10}**
N2	732-1091	TOS40-N-TO-2010 TOS49-N-TO-2010 TOS58-N-TO-2010 36to-N-TO-2010-C1 37to-N-TO-2010-C1 41to-N-TO-2010-C1 42to-N-TO-2010-C1 45to-N-TO-2010-C1 46to-N-TO-2010-C1 TOS39-N-TO-2010-C1 TOS48-N-TO-2010-C1 TOS90-N-TO-2010-C1	1pk-N-PK-2004 1429A-N-AU 1429B-N-AU 602-N-AU-1996 737-N-AU-1997 B2818-N-AU-2011 B2823-N-AU-2011 B2826-N-AU-2011 B2830-N-AU-2011 B2833-N-AU-2011 B2834-N-AU-2011 B2844-N-AU-2011 B2846-N-AU-2011 HE864320-N-PK KP8-N-AU-1989 All A1 1/1 19rw-N-RW-2009-C2 20rw-N-RW-2009-C2 26pk-N-PK-2004-C2 3in-N-IN-2007-C2 BU9-N-CD-2012-C2 Q553-N-LK-1995-C2 All C3 22/22	TOS14-N-TO-2010 TOS19-N-TO-2010 TOS4-N-TO-2010 TOS5-N-TO-2010 TOS53-N-TO-2010 TOS55-N-TO-2010 TOS57-N-TO-2010 TOS61-N-TO-2010 TOS63B-N-TO-2010 TOS64-N-TO-2010 TOS67-N-TO-2010 TOS68-N-TO-2010 TOS7-N-TO-2010 TOS71-N-TO-2010 TOS74-N-TO-2010 TOS78-N-TO-2010 TOS80-N-TO-2010 TOS82-N-TO-2010 TOS85-N-TO-2010 TOS86-N-TO-2010 TOS87-N-TO-2010 TOS88-N-TO-2010 TOS89-N-TO-2010 TOS93-N-TO-2010-C1 35to-N-TO-2010-C1 38to-N-TO-2010-C1 39to-N-TO-2010-C1	RGMCS**T**	**2.26x10^{-08}**

				40to-N-TO-2010-C1 43to-N-TO-2010-C1 536-N-TO-1993-C1 KP4-N-TO-1990-C1 Q276-N-TO-1989-C1 Q277-N-TO-1989-C1 Q278-N-TO-1989-C1 Q570-N-TO-1990-C1 TOS2-N-TO-2010-C1 TOS20-N-TO-2010-C1 TOS21-N-TO-2010-C1 TOS22-N-TO-2010-C1 TOS25-N-TO-2010-C1 TOS29-N-TO-2010-C1 TOS42-N-TO-2010-C1 TOS56-N-TO-2010-C1 TOS60-N-TO-2010-C1 TOS63A-N-TO-2010-C1 TOS65-N-TO-2010-C1 TOS91-N-TO-2010-C1		
N3	804-1229	ABTV2-N-PH	Unknown	ABTV1-N-MY	RMST	**5.26x10^{-22}**
N4	1167-1198	BU6-N-CD-2012-C2	ABTV1-N-MY	1pk-N-PK-2004 1429A-N-AU 1429B-N-AU 22in-N-IN 547-N-BI-1995 602-N-AU-1996 64in-N-IN-2009 737-N-AU-1997 B2818-N-AU-2011 B2823-N-AU-2011 B2826-N-AU-2011 B2830-N-AU-2011 B2833-N-AU-2011 B2834-N-AU-2011 B2844-N-AU-2011 B2846-N-AU-2011 HE864320-N-PK KP8-N-AU-1989 TOS4-N-TO-2010 TOS74-N-TO-2010 TOS78-N-TO-2010 TOS82-N-TO-2010 TOS88-N-TO-2010 All A1 1/1 4to-N-TO-2010-C1 45to-N-TO-2010-C1 46to-N-TO-2010-C1 TOS39-N-TO-2010-C1 TOS42-N-TO-2010-C1 TOS83-N-TO-2010-C1 TOS93-N-TO-2010-C1 All C2 except 3 *BU13-N-CD-2012-C2* *KP5-N-LK-2003-C2* *BU6-N-CD-2012-C2 (recombinant)* All C3 except 3 *482-96-N-AU-1996-C3* *482-97-N-AU-1997-C3* *482-98-N-AU-1998-C3*	RGB	**7.42x10^{-06}**
N6	1172-16	MP1-N-TW-1996 MP2-N-TW-1996-D6[#]	Unknown	Q280-N-WS-1989 768-N-PH-1995 Q568-3-N-ID-1995 625-N-TW-1996 626-N-TW-1996 523-6B-N-IN-1991-D8 All D5 except 1 *25tw-N-TW-D5*	RGB	**1.03x10^{-03}**
N7	1179-1202	TOS16-N-TO-2010-C1 TOS22-N-TO-2010-C1 TOS56-N-TO-2010-C1 TOS61-N-TO-2010 BU13-N-CD-2012-C2[#]	ABTV2-N-PH	TOS93-N-TO-2010-C1	RGB	**7.28x10^{-03}**

3.6. References

Abdel-Salam, S. M., Dahot, M. U. & Sadik, A. S. (2012). Molecular comparative analysis of component 1 (DNA-R) of an Egyptian isolate of banana bunchy top nanovirus isolated from banana aphid (Pentalonia nigronervosa). *Journal of Genetic Engineering and Biotechnology* 10, 55-65.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Amin, I., Qazi, J., Mansoor, S., Ilyas, M. & Briddon, R. W. (2008). Molecular characterisation of Banana bunchy top virus (BBTV) from Pakistan. *Virus Genes* 36, 191-198.

Anandhi, J., Vijila, C., Viswanath, G. S. & Lokeswari, T. S. (2007). Screening banana plants for banana bunchy top virus with primers specific to Indian isolates. *Journal of Plant Diseases and Protection* 114, 101-107.

Aronson, M. N., Meyer, A. D., Gyorgyey, J., Katul, L., Vetten, H. J., Gronenborn, B. & Timchenko, T. (2000a). Clink, a nanovirus-encoded protein, binds both pRB and SKP1. *J Virol* 74, 2967-2972.

Aronson, M. N., Meyer, A. D., Györgyey, J., Katul, L., Vetten, H. J., Gronenborn, B. & Timchenko, T. (2000b). Clink, a nanovirus-encoded protein, binds both pRB and SKP1. *J Virol* 74, 2967-2972.

Banerjee, A., Roy, S., Behere, G. T., Roy, S. S., Dutta, S. K. & Ngachan, S. V. (2014). Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. *Virus Res* 183, 41-49.

Bashir, R., Javed, F., Ahmed, R. & Mansoor, S. (2012). Use of rolling circle amplification for the identification of unknown components of banana bunchy top virus from Pakistan. *Pakistan Journal of Life and Social Sciences* 10, 91-97.

Bell, K. E., Dale, J. L., Ha, C. V., Vu, M. T. & Revill, P. A. (2002). Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Arch Virol* 147, 695-707.

Boni, M. F., Posada, D. & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176, 1035-1047.

Burns, T. M., Harding R. M. & Dale, J. L. (1995). The genome organisation of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76, 1471-1482.

Burns, T. M., Harding, R. M. & Dale, J. L. (1994). Evidence that banana bunchy top virus has a multiple component genome. *Arch Virol* 137, 371-380.

Denham, T. P., Haberle, S. G., Lentfer, C., Fullagar, R., Field, J., Therin, M., Porch, N. & Winsborough, B. (2003). Origins of agriculture at Kuk Swamp in the highlands of New Guinea. *Science* 301, 189-193.

Duay, J. A. M., Miller, R. H., Wall, G. C., Pike, K. S. & Foottit, R. G. (2014). Pentalonia nigronervosa Coquerel and Pentalonia caladii van der Goot (Hemiptera: Aphididae) and their relationship to Banana bunchy top virus in Micronesia. *Pac Sci* 68, 359-364.

Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus. *J Virol* 82, 957-965.

Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9, 267-276.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-1797.

FAOSTAT Food and Agriculture Organisation of the United Nations Statistics Division. 2012.

Firth, C., Charleston, M. A., Duffy, S., Shapiro, B. & Holmes, E. C. (2009). Insights into the evolutionary history of an emerging livestock pathogen: Porcine circovirus 2. *J Virol* 83, 12813-12821.

Foottit, R. G., Maw, H. E. L., Pike, K. S. & Miller, R. H. (2010). The identity of Pentalonia nigronervosa Coquerel and P. caladii van der Goot (Hemiptera: Aphididae) based on molecular and morphometric analysis. *Zootaxa*, 25-38.

Fu, H. C., Hu, J. M., Hung, T. H., Su, H. J. & Yeh, H. H. (2009). Unusual events involved in banana bunchy top virus strain evolution. *Phytopathology* 99, 812-822.
Furuya, N., Dizon, T. O. & Natsuaki, K. T. (2006). Molecular Characterization of Banana bunchy top virus and Cucumber mosaic virus from Abaca (Musa textilis Nee). *Journal of Agricultural Science-Tokyo University of Agriculture* 51, 92-101.
Furuya, N., Kawano, S. & Natsuaki, K. T. (2005). Characterization and genetic status of banana bunchy top virus isolated from Okinawa, Japan. *J Gen Plant Pathol* 71, 68-73.
Furuya, N., Somowiyarjo, S. & Natsuaki, K. (2004). Virus detection from local banana cultivars and the first molecular characterization of Banana bunchy top virus in Indonesia. *Journal of Agricultural Science-Tokyo University of Agriculture* 49, 75-81.
Geering, A. D. W. & Thomas, J. E. (1997). Search for alternative hosts of banana bunchy top virus in Australia. *Australas Plant Pathol* 26, 250-254.
Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573-582.
Gorbalenya, A. E., Koonin, E. V. & Wolf, Y. I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett* 262, 145-148.
Grigoras, I., Ginzo, A. I. d. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.-J. & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J Gen Virol* 95, 1178-1191.
Grigoras, I., Timchenko, T., Grande-Perez, A., Katul, L., Vetten, H. J. & Gronenborn, B. (2010). High variability and rapid evolution of a nanovirus. *J Virol* 84, 9105-9117.
Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59, 307-321.
Hafner, G. J., Harding, R. M. & Dale, J. L. (1997a). A DNA primer associated with banana bunchy top virus. *J Gen Virol* 78, 479-486.
Hafner, G. J., Stafford, M. R., Wolter, L. C., Harding, R. M. & Dale, J. L. (1997b). Nicking and joining activity of banana bunchy top virus replication protein in vitro. *The Journal of general virology* 78 (Pt 7), 1795-1799.
Harding, R. M., Burns, T. M. & Dale, J. L. (1991). Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *J Gen Virol* 72, 225-230.
Harding, R. M., Burns, T. M., Hafner, G., Dietzgen, R. G. & Dale, J. L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *J Gen Virol* 74, 323-328.
Harkins, G. W., Delport, W., Duffy, S., Wood, N., Monjane, A. L., Owor, B. E., Donaldson, L., Saumtally, S., Triton, G., Briddon, R. W., Shepherd, D. N., Rybicki, E. P., Martin, D. P. & Varsani, A. (2009). Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol J* 6, 104.
Harkins, G. W., Martin, D. P., Christoffels, A. & Varsani, A. (2014). Towards inferring the global movement of beak and feather disease virus. *Virology* 450-451, 24-33.
He, Z. F., Li, H. P., Xiao, H. G. & Fan, H. Z. (2000). Cloning and sequencing of DNA component 1 of two BBTV strains. *Chih Wu Ping Li Hsueh Pao* 30, 364-369.
He, Z. F., Li, H. P., Xiao, H. G. & Fan, H. Z. (2001a). Cloning and sequencing of DNA component 3 of two BBTV strains. *Nongye Shengwu Jisu Xuebao* 9, 145-148.
He, Z. F., Li, H. P., Xiao, H. G. & Fan, H. Z. (2001b). Cloning and sequencing of DNA component 6 of two BBTV strains. *Hua Nan Nung Yeh Ta Hsueh Hsueh Pao* 22, 33-36.
Herrera-Valencia, V. A., Dugdale, B., Harding, R. M. & Dale, J. L. (2006). An iterated sequence in the genome of banana bunchy top virus is essential for efficient replication. *J Gen Virol* 87, 3409-3412.
Hooks, C. R. R., Wright, M. G., Kabasawa, D. S., Manandhar, R. & Almeida, R. P. P. (2008). Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. *Annuals of Applied Biology* 153, 1-9.
Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. (2007). Reassortment and concerted evolution in Banana bunchy top virus genomes. *J Virol* 81, 1746-1761.

Hu, J. S., Wang, M., Sether, D., Xie, W. & Leonhardt, K. W. (1996). Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (Pentalonia nigronervosa). *Ann Appl Biol* 128, 55-64.
Hyder, M. Z. (2009). Sequencing and Genetic Characterization of major components of Banana Bunchy Top Virus. In *Department of Biochemistry*. Pakistan: Arid Agriculture University.
Hyder, M. Z., Raza, S. Q., Hameed, S., Khalid, S. & Saqlan Naqvi, S. M. (2007). Phylogenetic relationship of TJ1 isolate of Banana bunchy top virus from Pakistan by DNA-R sequence analysis. *Can J Plant Pathol* 29, 63-68.
Hyder, M. Z., Shah, S. H., Hameed, S. & Naqvi, S. M. S. (2011). Evidence of recombination in the Banana bunchy top virus genome. *Infect Genet Evol* 11, 1293-1300.
Islam, M. N., Naqvi, A. R., Jan, A. T. & Haq, Q. M. R. (2010a). Genetic Diversity and Possible Evidence of Recombination among Banana Bunchy Top Virus (BBTV) Isolates. *International Research Journal of Microbiology* Vol 1, 001-012.
Islam, M. N., Naqvi, A. R., Jan, A. T. & Haq, Q. M. R. (2010b). Genetic dversity and possible evidence of recombination among Banana bunchy top virus (BBTV) isolates. *International Research Journal of Microbiology* Vol 1, 001-012.
Izquierdo-Carrasco, F., Smith, S. A. & Stamatakis, A. (2011). Algorithms, data structures, and numerics for likelihood-based phylogenetic inference of huge trees. *BMC bioinformatics* 12, 470.
James, A. P. (2011). Viruses of Banana in East Africa: Queensland University of Technology.
Jun, T. E. Z. & Zhi-Xin, L. (2005). Cloning and sequencing of DNA components of Banana bunchy top virus Hainan isolate. *Chinese Journal of Agricultural Biotechnology* 2, 91-97.
Karan, M., Harding, R. M. & Dale, J. L. (1994). Evidence for two groups of banana bunchy top virus isolates. *J Gen Virol* 75, 3541-3546.
King, A. M., Lefkowitz, E., Adams, M. J. & Carstens, E. B. (2012). Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses. Burlington: Elsevier Science.
Kraberger, S., Harkins, G. W., Kumari, S. G., Thomas, J. E., Schwinghamer, M. W., Sharman, M., Collings, D. A., Briddon, R. W., Martin, D. P. & Varsani, A. (2013). Evidence that dicot-infecting mastreviruses are particularly prone to inter-species recombination and have likely been circulating in Australia for longer than in Africa and the Middle East. *Virology* 444, 282-291.
Kumar, P., Hanna, R., Alabi, O., Soko, M., Oben, T., Vangu, G. & Naidu, R. (2011). Banana bunchy top virus in sub-Saharan Africa: investigations on virus distribution and diversity. *Virus Res* 159, 171.
La, P., Cai, W. Q. & Fang, R. X. (2000). Cloning and sequencing of banana bunchy top virus genome component I. *Chinese Journal of Virology* 16, 158-161.
Lageix, S., Catrice, O., Deragon, J. M., Gronenborn, B., Pélissier, T. & Ramirez, B. C. (2007). The nanovirus-encoded clink protein affects plant cell cycle regulation through interaction with the retinoblastoma-related protein. *J Virol* 81, 4177-4185.
Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009). Widely Conserved Recombination Patterns among Single-Stranded DNA Viruses. *J Virol* 83, 2697-2707.
Lejju, B. J., Robertshaw, P. & Taylor, D. (2006). Africa's earliest bananas? *J Archaeol Sci* 33, 102-113.
Magee, C. J. P. (1927). Investigation on the bunchy top disease of the banana. *Bulletin of the Council for Scientific and Industrial Research in Australia* 30.
Magee, C. J. P. (1940). Transmission studies on the Banana Bunchy-top Virus. *Journal of the Australian Institute of Agricultural Science* 6, 109-110.
Mandal, B., Shilpi, S., Barman, A. R., Mandal, S. & Varma, A. (2013). Nine novel DNA components associated with the foorkey disease of large cardamom: evidence of a distinct babuvirus species in Nanoviridae. *Virus Res* 178, 297-305.
Manickam, K., Doraiswamy, S., Ganapathy, T. & Rabindran, R. (2002). Early detection of banana bunchy top virus in India using polymerase chain reaction. *Acta Phytopathol Entomol Hung* 37, 9-16.
Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562-563.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462-2463.
Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrovir* 21, 98-102.
Muhire, B. M., Varsani, A. & Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* 9, e108277.
Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L. & Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol Biol Evol* 30, 1196-1205.
Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K. & Pond, S. L. K. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS genetics* 8, e1002764.
Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., Martin, D. P. & Varsani, A. (2007). Successful application of FTA® Classic Card technology and use of bacteriophage phi 29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *J Virol Methods* 140, 100-105.
Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218-225.
Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J. P., Jenny, C., Lebot, V., Risterucci, A. M., Tomekpe, K., Doutrelepont, H., Ball, T., Manwaring, J., de Maret, P. & Denham, T. (2011). Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc Natl Acad Sci U S A* 108, 11311-11318.
Pinili, M. S., Nagashima, I., Dizon, T. O. & Natsuaki, K. T. (2013). Cross-transmission and new alternate hosts of Banana bunchy top virus. *Tropical Agriculture and Development* 57, 1-7.
Pinili, M. S., Nyana, D., Suastika, G. & Natsuaki, K. (2011). Molecular analysis of Banana bunchy top virus first isolated in Bali, Indonesia. *J Agric Sci Tokyo Univ Agric* 56, 125-134.
Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* 82, 655-665.
Pond, S. L. K., Frost, S. D. W. & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676-679.
Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25, 1253-1256.
Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci U S A* 98, 13757-13762.
Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* 54, 396-402.
Ram, R. D. & Summanwar, A. S. (1984). Colocasia esculenta (L.) schott. A reservoir of bunchy top disease of banana. *Current Science* 53, 145-146.
Randrianja, S. & Ellis, S. (2009). *Madagascar: A Short History*: University of Chicago Press.
Rosario, K., Duffy, S. & Breitbart, M. (2012). A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch Virol* 157, 1851-1871.
Savory, F. R. & Ramakrishnan, U. (2014). Asymmetric patterns of reassortment and concerted evolution in Cardamom bushy dwarf virus. *Infection, Genetics and Evolution* 24, 15-24.
Schierup, M. H. & Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156, 879-891.
Selvarajan, R., Sheeba, M. M., Balasubramanian, V., Rajmohan, R., Dhevi, N. L. & Sasireka, T. (2010). Molecular Characterization of Geographically Different Banana bunchy top virus Isolates in India. *Indian Journal of Virology* 21, 110-116.
Sharman, M., Thomas, J. E., Skabo, S. & Holton, T. A. (2008a). Abaca bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch Virol* 153, 135-147.
Sharman, M., Thomas, J. E., Skabo, S. & Holton, T. A. (2008b). Abaca bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch Virol* 153, 135-147.
Shepherd, D. N., Martin, D. P., Lefeuvre, P., Monjane, A. L., Owor, B. E., Rybicki, E. P. & Varsani, A. (2008). A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *J Virol Methods* 149, 97-102.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* 34, 126-129.
Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* 93, 1103-1119.
Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
Su, H.-J., Wu, R.-Y. & Tsao, L.-Y. (1992). Ecology of banana bunchy-top virus disease. In *Proceedings of the International Symposium on Recent Developments in Banana Cultivation Technology, Los Baños, Philippines: INIBAP-ASPNET*, pp. 308-312.
Su, H. J., Tsao, L. Y., Wu, M. L. & Hung, T. H. (2003). Biological and molecular categorization of strains of Banana bunchy top virus. *Journal of Phytopathology* 151, 290-296.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.
Thomas, J. E. & Dietzgen, R. G. (1991). Purification, characterization and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *J Gen Virol* 72, 217-224.
van der Walt, E., Martin, D. P., Varsani, A., Polston, J. E. & Rybicki, E. P. (2008). Experimental observations of rapid Maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virology Journal* 5, 104.
Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J., Murilo Zerbini, F. & Martin, D. (2014). Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch Virol*, 1-11.
Vega-Rocha, S., Gronenborn, B., Gronenborn, A. M. & Campos-Olivas, R. (2007). Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circoviras structures. *Biochemistry* 46, 6201-6212.
Vishnoi, R., Raj, S. & Prasad, V. (2009). Molecular characterization of an Indian isolate of Banana bunchy top virus based on six genomic DNA components. *Virus Genes* 38, 334-344.
Wang, H. I., Chang, C. H., Lin, P. H., Fu, H. C., Tang, C. & Yeh, H. H. (2013). Application of Motif-Based Tools on Evolutionary Analysis of Multipartite Single-Stranded DNA Viruses. *PLoS ONE* 8.
Wanitchakorn, R., Hafner, G. J., Harding, R. M. & Dale, J. L. (2000a). Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *The Journal of general virology* 81, 299-306.
Wanitchakorn, R., Harding, R. M. & Dale, J. L. (2000b). Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch Virol* 145, 593-602.
Wanitchakorn, R., Harding, R. M. & Dale, J. L. (2000c). Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch Virol* 145, 593-602.
Xie, W. S. & Hu, J. S. (1995). Molecular cloning, sequence analysis, and detection of banana bunchy top virus in Hawaii. *Phytopathology* 85, 339-347.
Yu, N. T., Zhang, Y. L., Feng, T. C., Wang, J. H., Kulye, M., Yang, W. J., Lin, Z. S., Xiong, Z. & Liu, Z. X. (2012). Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* 44, 488-494.

Chapter 4

Defective molecules of *Banana bunchy top virus* and the characterisation of the common regions found in babuviruses and alphasatellite molecules

---

Contents

This work is currently under review with Archives of Virology and is presented in a similar manner to that of the manuscript:

Stainton, D., Martin, D.P., Thomas, J.E., Varsani, A. Defective molecules of *Banana bunchy top virus* and the characterisation of the common regions found in babuviruses and alphasatellite molecules. (In review)

4.1 Abstract

Banana bunchy top virus (BBTV; *Babuvirus* genus), like other members of the *Nanoviridae,* is a multi-component single-stranded DNA virus. Each of the six components comprising the BBTV genome contain two common regions: the common region stem-loop (CR-SL) and the common region major (CR-M). For all components the CR-SL is involved in the initiation of rolling circle replication (RCR) whereas the CR-M is involved in secondary strand synthesis. The CR-SL is highly conserved across all three babuvirus species. In addition to the six integral BBTV genome components, alphasatellite molecules have been found associated with some BBTV infections as well as those of other nanoviruses and geminiviruses. These molecules contain a single open reading frame (ORF) encoding a replication-associated protein (Rep). A sudden increase in the number of publically available babuvirus sequences, has recently provided the opportunity for the in-depth analysis of common regions in babuviruses and their alphasatellites. Besides the available full genome component sequences, twenty-four apparently 'defective' BBTV components have also been identified within BBTV infected plants. All of the identified defective molecules have retained at least part of the CR-SL and CR-M but have insertions and/or deletions that in most cases result in open reading frame disruptions. Interestingly, 23/24 of these defective molecules were apparently derived from DNA-R. Two breakpoint hotspot regions were identified for the defective DNA-R and two common lineages were also identified.

4.2 Introduction

Babuviruses are single-stranded DNA viruses which infect monocotyledonous plants. There are three accepted babuviruses: BBTV, *Abaca bunchy top virus* (ABTV) and *Cardamom bushy dwarf virus* (CBDV). Both BBTV and ABTV infect members of the Musaceae family whereas CBDV infects large cardamom. Babuviruses are transmitted by aphids in the *Pentalonia* genus (Basu & Ganguly, 1968; Blackman & Eastop, 2008; Magee, 1927; Mandal *et al.*, 2004; Varma & Capoor, 1964; Watanabe *et al.*, 2013), with symptoms of infection including plant stunting, leaf bunching and yield loss. A single sequence has been identified which is associated with Coconut foliar decay disease (Rohde *et al.*, 1990). This isolate contains a Rep-like gene and has been classified as both an unassigned Babuvirus and an alphasatellite (King *et al.*, 2011), here this sequences is referred to as Coconut foliar decay alphasatellite (CFDA) and it is analysed as an alphasatellite.

All babuviruses have a multi-component genome with six individually encapsidated components which are all considered indispensible for the initiation of infection: DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C, and DNA-N. CBDV contains two additional components, DNA-Uf1 and DNA-Uf2, neither of which contain obvious open reading frame/s (ORFs), but which might still be required for the initiation of CBDV infections (Mandal *et al.*, 2013). DNA-R, contains both a gene that encodes a replication-associated protein (Rep), and a second ORF of unknown function that is transcribed but might not be translated; DNA-S contains a gene encoding the viral capsid protein (CP); DNA-M contains a gene that encodes a movement protein (MP); DNA-C contains a gene that encodes a cell-cycle link protein (Clink); DNA-N contains a gene that encodes a nuclear shuttle protein (NSP); and DNA-U3, contains no obvious ORFs and has no known function (Beetham *et al.*, 1997; Burns *et al.*, 1995; King *et al.*, 2011; Mandal *et al.*, 2013; Sharman *et al.*, 2008). All canonical components of the babuviruses contain two common regions which are highly conserved amongst the components of each species: the common region stem-loop (CR-SL) and the common region major (CR-M) (Burns *et al.*, 1995; Mandal *et al.*, 2013; Sharman *et al.*, 2008).

A number of additional DNA molecules are sometimes associated with BBTV under field infections. These BBTV-associated DNA molecules include DNA-R-like molecules, called alphasatellites (formerly known as satellites) that encode a Rep, and apparently defective genome components that contain insertions and/or deletions (unpublished sequences on

GenBank) (Bell *et al.*, 2002; Fu *et al.*, 2009; Horser *et al.*, 2001b; Wu *et al.*, 1994; Yu *et al.*, 2012). Alphasatellites have also been found associated with other nanoviruses and geminiviruses (Grigoras *et al.*, 2014; Katul *et al.*, 1998; King *et al.*, 2011; Saunders & Stanley, 1999; Zhou, 2013) and are ~20% larger in size than nanovirus genome components and ~50% smaller than geminivirus genome components. Although they are able to replicate on their own within an infected cell, alphasatellites are unable to trans-replicate the genome components of nanoviruses and geminiviruses (Horser *et al.*, 2001a; Timchenko *et al.*, 1999; Timchenko *et al.*, 2000), and must be encapsidated by capsid proteins encoded by these associated viruses in order to be transmitted (Briddon & Stanley, 2006).

Since babuviruses have only one Rep encoding component, all five other components, and presumably also the defective genome components, need to be trans-replicated by this Rep and must therefore possess Rep recognition sequences near their origins of virion strand replication (*v-oris*). This is achieved via the CR-SL which contains three repeated (or iterated) five nucleotide (nt) long sequence elements (called iterons R, F1 and F2) and a hairpin-loop sequence containing a conserved nonanucleotide motif "TATTATTAC". These sequence features are involved in the recognition of the *v-ori* by Rep and the initiation by Rep of rolling circle replication (RCR) (Hafner *et al.*, 1997b; Herrera-Valencia *et al.*, 2006). Whereas Rep recognises and possibly binds to the CR-SL at or near the iterons, it initiates RCR when it nicks a bound ssDNA viral component sequence between nucleotides seven and eight of the nonanucleotide motif (Hafner *et al.*, 1997b), (Herrera-Valencia *et al.*, 2006). In BBTV and ABTV iteron R (GTCCC) is 5' of the *v-ori* hairpin sequence, whereas iterons F1 (GGGAC) and F2 (GGGAC) are 3' of the hairpin. The arrangement of iterons in CBDV is the same as the sequences of the iterons R and F1, however, the F2 sequence is slightly different (GGAAC) (Burns *et al.*, 1995; Mandal *et al.*, 2013; Sharman *et al.*, 2008). Although mutations in all of the iterons impact BBTV replication, it is apparent that the F2 iteron is the most crucial (Herrera-Valencia *et al.*, 2006).

The majority of studies of BBTV have suggested that BBTV DNA-U3 is apparently unique as it lacks an iteron R and contains a smaller CR-SL which starts 14-21 nt closer to the stem-loop than that of the other components (Banerjee *et al.*, 2014; Burns *et al.*, 1995; Islam *et al.*, 2010; Vishnoi *et al.*, 2009). However, it has also been suggested that DNA-U3 might in fact have a larger CR-SL with an iteron R sequence ~ 90bp further downstream from the stem-loop than in other components (Herrera-Valencia *et al.*, 2006). Using a motif based tool,

Wang *et al.* (2013) also identified a region which was present within the CR-SL of all other components but occurred further downstream of the CR-SL in DNA-U3. The potential dispensability of iteron R within the CR-SL of some babuvirus genome components is further evident as the CR-SL of the CBDV DNA-N component is also apparently missing (Mandal *et al.*, 2013). The presence of iteron R within the CR-SL of babuviruses is, however, almost certainly selectively favoured during evolution since all other components of BBTV, ABTV and CBDV contain all three iterons (Burns *et al.*, 1995; Mandal *et al.*, 2013; Sharman *et al.*, 2008) and mutations within iteron R of the BBTV DNA-N component have been shown to have a greater impact on DNA-N replication than similar mutations in iteron F1 (Herrera-Valencia *et al.*, 2006).

Horser *et al.* (2001b) identified the CR-SL of the BBTV alphasatellites but only found obvious similarities between the BBTV genome components and BBTV associated alphasatellites within the hairpin structures at their *v-oris*. Most notably, both had similar nonanucleotide sequences - TAGTATTAC for the alphasatellites and TATTATTAC for the BBTV components. In addition to BBTV alphasatellites, a CBDV-associated alphasatellite has been identified (Mandal *et al.*, 2013), but to date no alphasatellites have been identified associated with ABTV infections. The distance between the CR-SL and CR-M is variable from component to component. The CR-M is involved in promoting transcription (Burns *et al.*, 1995) and is also the binding site of a primer which is involved in the initiation of complementary strand replication (Hafner *et al.*, 1997a). CBDV has two additional genome components CBDV-Uf1 and Uf2, both of which do not have any obvious ORFs but contain CR-M and CR-SL regions similar to those of the integral components (Mandal *et al.*, 2013). Here we further characterise the CR-SL and CR-M sequences of babuviruses and their associated satellites. We attempt to identify the CR-SL iteron R in BBTV DNA-U3 and CBDV DNA-N and re-evaluate the relationships between these regions and the known babuvirus alphasatellite sequences.

4.3 Materials and Methods

4.3.1 Common regions of canonical BBTV, ABTV and CBDV component datasets

Separate component-specific datasets of BBTV, ABTV (Chapter Two and Chapter Three), and CBDV (downloaded from GenBank 10/07/14), were aligned with MUSCLE using default settings (Edgar, 2004) implemented in MEGA 5 (Tamura *et al.*, 2011). CR-M and CR-SL were identified for each component (including for the DNA-Uf1 and DNA-Uf2 components of CBDV). For each species and component the common region sequences were separated into CR-M and CR-SL datasets. All accession numbers for BBTV, ABTV and CBDV sequences are present in Supplementary Table 4.1.

Given that the CR-SL of BBTV DNA-U3 and of CBDV DNA-N had previously been identified without an iteron R, a region 5' of the previously identified CR-SL boundary was also included in the CR-SL datasets. The CR-SL sequences of all three babuvirus species were aligned together using MAFFT with the G-INS-I setting (Katoh & Standley, 2013). Based on this alignment the 5' region of the alignment was trimmed to include only CR-SL sequences that were obviously homologous between different components and the remaining sequence fragments were realigned with MAFFT with the G-INS-I setting. The refined CR-SL of all components was re-split into species- and component-specific datasets and a consensus representation of each of these datasets was constructed using WebLogo v3.4 (Crooks *et al.*, 2004). The BBTV DNA-U3 CR-SL dataset had a large number of sequences which contained an insert within the CR-SL region and this dataset was therefore split into two separate datasets, such that two separate CR-SL sequence logos were generated for BBTV DNA-U3. Sequence Demarcation Tool (SDT) v1.2 (Muhire *et al.*, 2014) was used to determine the pairwise identities of the iterons R, F1 and F2.

An attempt was made to align the CR-M dataset which contained the CR-M sequences of all components of all three babuvirus species. However, due to low degrees of sequence similarity between the sequences drawn from the different species, the CR-M of each species was eventually aligned separately in MAFFT with the G-INS-I setting. For each species, the percentage pairwise nucleotide identities of the CR-M of all components, were determined in SDT v1.2 with the MUSCLE alignment option (Muhire *et al.*, 2014).

These three alignments were re-split and a consensus representation of the CR-M sequences for each component for each species was then generated using Weblogo v3.4. From the

sequence logo, a portion of CR-M displaying detectable degrees of sequence similarity across all three species was visually identified. This highly conserved portion of the CR-M sequences contained the GC-rich region and this portion was realigned using MAFFT with the G-INS-I setting with sequence logos generated as above.

4.3.2 Common regions of alphasatellites

The non-coding regions of the alphasatellites (downloaded 30/09/14) were first aligned with MUSCLE with default settings (Edgar, 2004) implemented in MEGA 5 (Tamura *et al.*, 2011). Regions containing the nonanucleotide were identified and further aligned to identify the CR-SL. The alphasatellite CR-SL, and the CR-SL sequences of BBTV, ABTV and CBDV DNA-R components were aligned using MAFFT with the G-INS-I setting. This alignment was then separated into five groups consisting of (1) DNA-R CR-SL sequences; (2) babuvirus alphasatellites (babu-alphas) CR-SL sequences, (3) nanovirus alphasatellite (nano-alphas) CR-SL sequences, (4) geminiviruses alphasatellite (gemini-alphas) CR-SL sequences; and (5) a CFDA sequence. Weblogo v3.4 was used to generate sequence logos of each of the five groups.

The second common region, the CR-M, was identified by realigning GC-rich regions of the various DNA-R and alphasatellite sequences. Once identified, the probable CR-M sequences, were aligned using MAFFT together with the DNA-R CR-M sequences of ABTV, BBTV and CBDV. However, as the CR-M sequences of the alphasatellites were highly diverse even within the five major groups, the CR-M sequences of the five different groups were separately realigned with MAFFT with the G-INS-I setting. Some of these groups were further split, with the gemini-alpha group being split into two groups and the babu-alpha group being split into four groups (BBTV-1, BBTV-2, CBDV and a group containing all Babu-alphas). The percentage pairwise identities for the CR-M of each group of alphasatellites was determined by SDT v1.3 with MUSCLE alignment (Muhire *et al.*, 2014). Accession numbers of the alphasatellites sequences present in each group, for both the CR-M and CR-SL are present in Supplementary Table 4.2. Sequence logos of the CR-Ms were generated using Weblogo as above. The CR-M of the nano-alphas was not obviously conserved, and therefore a sequence logo was not generated for this group. It is also noteworthy that we were unable to identify both a CR-SL and CR-M for every alphasatellite sequence which was examined.

4.3.3 Defective molecules

The sequences of defective BBTV genome components were determined during attempts to clone and sequence full BBTV genome sequences from BBTV infected banana plants, using previously designed BBTV specific primers (Stainton *et al.*, 2012; Stainton *et al.*, 2015). These back-to-back primers also opportunistically amplified a number of defective molecules. The amplicons were then cloned into pGEM T-easy vector (Promega, USA) and the resulting clones were Sanger-sequenced at Macrogen Inc. (Korea). BBTV component alignments were used to identify defective components. The defective DNA molecules were then each aligned individually with a reference sequence (TOS60;(Stainton *et al.*, 2015)) of either component DNA-R (KM607707) or DNA-U3 (KM607835) in order to identify insertions and deletions relative to the architecture of a full component. Insert regions in the defective molecules were compared against the full component to identify repeat regions. Additionally, BLASTn (Altschul *et al.*, 1990) analysis was used to identify potential recombination regions derived from other components.

A maximum likelihood phylogenetic tree was constructed of the defective DNA-R molecules with detected insert regions removed and recombination free DNA-R sequences (DNA-R recombination free dataset from Chapter 3) (Stainton *et al.*, 2015). Sequences were first aligned in MUSCLE (Edgar, 2004), and a maximum phylogenetic tree was constructed with PhyML 3 (Guindon *et al.*, 2010) with best fit model determined using jModelTest (Posada, 2008) and 100 bootstrap replicates for branch support, implemented in MEGA5 (Tamura *et al.*, 2011). ABTV DNA-R sequences were used to root the phylogenetic tree with branches with bootstrap support of < 60% collapsed in Mesquite v2.75 (http://mesquiteproject.org/).

4.4 Results and Discussion

4.4.1 Evidence of variability within babuvirus CR-SL iteron sequences

The CR-SL were identified in all publically available BBTV, CBDV and ABTV genome component sequences and were found to be highly similar across the species (Figure 4.1). All components have regions which resemble iteron R, iteron F1 and iteron F2 including those of BBTV DNA-U3 and CBDV DNA-N. A greater degree of sequence conservation was observed for iteron R (with 88% of the iteron sequences sharing 100% pairwise identity) than either iteron F1 or F2 (Table 4.1). Additionally, the vast majority of the components of all three species possess an identical nonanucleotide, TATTATTAC, at their presumed *v-ori*s.

We noted two inserts within the CR-SL regions of BBTV DNA-U3 sequences, which affect the alignments of all components. One of these insertions, is due to two sequences sampled in Taiwan which contain an insert of either 16 or 17 nt in size and 3' of the *v-ori* hairpin loop, when aligned these result in an overlapping 22 nt region in the DNA-U3 of BBTV (vertical lines in Figure 4.1 between positions 94-115). The second insertion, of ~30 nt is 5' of the *v-ori* hairpin loop within 111 DNA-U3 sequences. This insert replaces a 9 nt section seen in the majority of sequences and results in an alignment gap of ~20 nt in all sequences. Seventy six of the 77 BBTV DNA-U3 sequences without this latter insertion have the expected GTCCC iteron R sequence. In contrast, none of the 111 sequences with the insertion have the iteron R (GTCCC) which is seen in majority of other components. Instead 95% of these 111 sequences have a GCCTC motif which may be functionally equivalent to iteron R. There is, however, a GTCCC motif ~50 nt 5' of this GCCTC motif which has previously been identified as the probable DNA-U3 iteron R (Herrera-Valencia *et al.*, 2006). But, based on the alignment of the CR-SL with all the components it is much more likely that GCCTC is acting as the iteron R for these DNA-U3 sequences. Therefore the full CR-SL and all iterons of BBTV DNA-U3 have been identified.

It is also noteworthy that as a consequence of using an alignment which encompassed a larger section of the DNA-N sequence than previous analysed (Mandal *et al.*, 2013) we were able to identify a probable iteron R sequence within the CR-SL of this component in CBDV (Figure 4.1).

Trans-replication has been shown experimentally with three nanovirus species, *Faba bean necrotic yellows virus* (FBNYV), *Milk vetch dwarf virus* (MDV) and *Subterranean clover stunt virus* (SCSV), where each DNA-R was able to replicate the DNA-C of the other species (Timchenko *et al.*, 2000). Similarities in the CR-SL of all three species suggest that this may also be true of the babuviruses. The two additional largely uncharacterised components of CBDV, CBDV-Uf1 (KF435145) and CBDV-Uf2 (KF435146) contain CR-SL sequences which are more similar to those of the six canonical CBDV components than they are to those of known babuvirus alphasatellites suggesting these are potentially genuine CBDV genome components.

Although the consensus iteron R of all six canonical babuvirus genome components is "GTCCC" a number of individual sequences also had variations of this consensus (GTCTC / GTCGC / GCCCC / TTCGC / CTCCC / CCCCC / GTCCT / ATCCC / CTCTC / TGCTC /

TCGCC / TCCCTC). There was, however, detectably more variation from the consensus iteron F1 (GGGAC) and iteron F2 (GGGAC) with only 48% and 44%, respectively, of the sequences sharing 100% identity (Table 4.1). Variability was found both between species (with iteron F1 in DNA-M for example AGGAC / GGAAC / AGAAC in BBTV, ABTV, CBDV respectively), and within species (with isolates of CBDV iteron F2 for example containing GGA/GAC) (Figure 4.1). Although complete mutagenesis of each of these three conserved iterons is known to result in varying degrees of replicative fitness loss (Herrera-Valencia *et al.*, 2006), the number of nucleotide variations which are seen within these iteron sequences suggests that there may exist a degree of flexibility in the exact nucleotide sequence motifs that Rep is able to recognise.

Table 4.1: Percentage pairwise identities of the iteron sequences, R, F1 and F2 identified using the sequence demarcation tool (SDT) v 1.2 (Muhire *et al.*, 2014).

Iteron R	
Proportion of pairs	% identity
8.81×10^{-1}	100
9.56×10^{-2}	60
2.05×10^{-2}	80
2.90×10^{-3}	40
9.20×10^{-6}	20

Iteron F1	
Proportion of pairs	% identity
4.80×10^{-1}	100
4.20×10^{-1}	80
9.87×10^{-2}	60
2.27×10^{-3}	40
2.91×10^{-6}	20

Iteron F2	
Proportion of pairs	% identity
4.54×10^{-1}	86
4.40×10^{-1}	100
1.02×10^{-1}	71
1.77×10^{-3}	83
6.13×10^{-4}	87
5.90×10^{-4}	43
4.67×10^{-4}	67
5.61×10^{-5}	29
5.11×10^{-5}	50
8.31×10^{-7}	33
4.16×10^{-7}	80

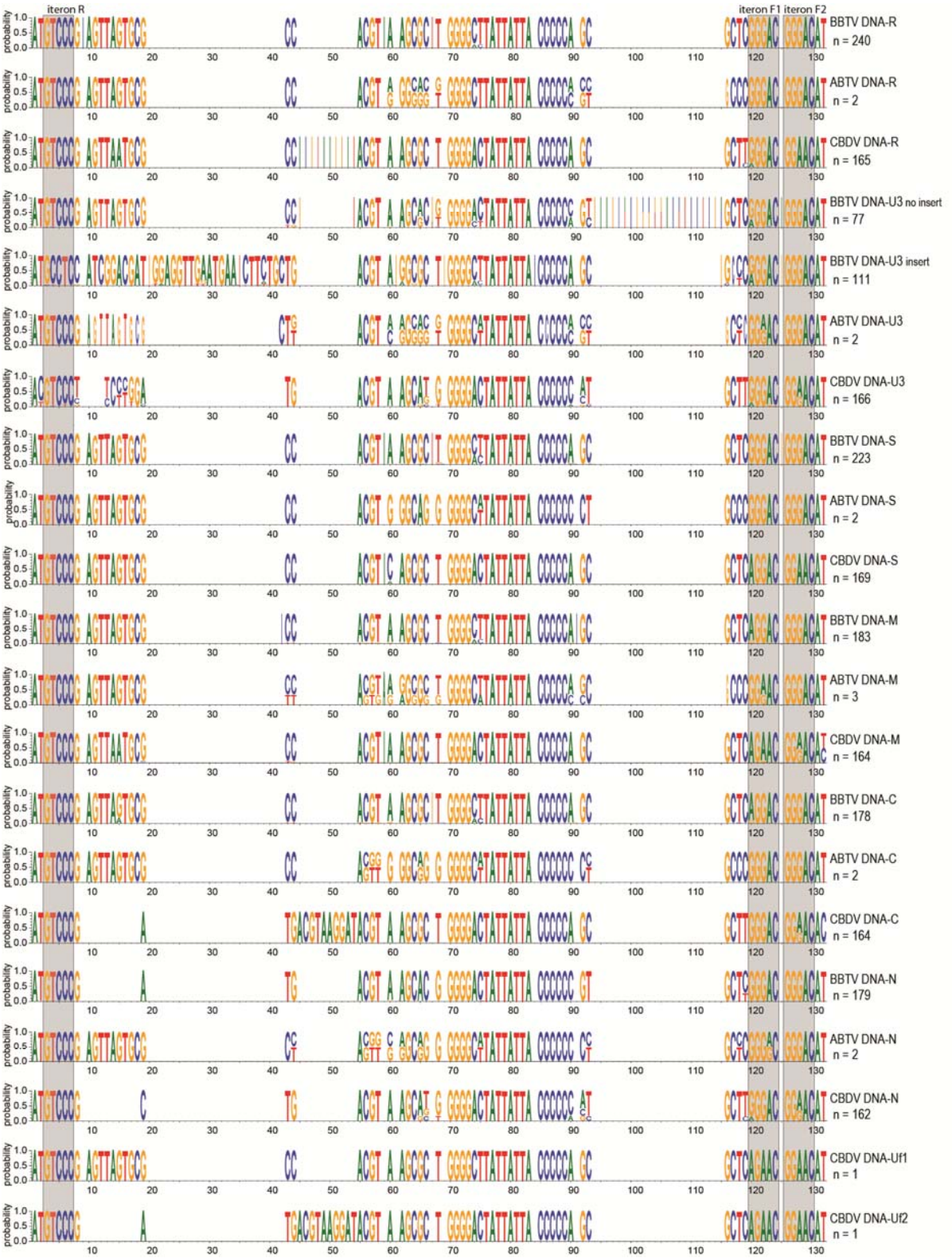


Figure 4.1: Aligned CR-SL sequences of canonical babuvirus genome components. Iteron R, F1 and F2 are highlighted in grey. The height of the letters indicates the relative proportions of each nucleotide at each site. The widths of the letters take into consideration the number of gaps at that site across all the isolates in the alignment, for instance when there are a large number of gaps at a specific site relative to the number of nucleotides, the width of the letter is narrow.

4.4.2 Evidence of a highly conserved GC-rich sequence within babuvirus CR-M sequences

The CR-M was identified in all three babuvirus species. All components of each species were individually aligned (Figure 4.2) and within each species all components had CR-M sequences that were >50% similar. However, unlike with the CR-SL sequences, we were unable to generate a credible alignment of CR-M sequences drawn from the different babuvirus species. Despite this, the 3' section of the CR-M of all species had a similar GC rich region. Therefore the ~50 nt portion at the 3' region of the CR-M containing this GC rich region of sequences from the different species were realigned to reveal a partially conserved region containing GGGCCGNAGGCCC sequence that was present in most of the available babuvirus genome component sequences (Figure 4.3). These GC-rich sequences are similar to the GC- boxes in the transcription-promoting rightward promoter element (Rpe1) found in geminiviruses (Burns et al., 1995; Fenoll et al., 1990). Interestingly, other GC-rich regions are also present in some single-stranded bacterial plasmids where they are involved in the priming of complementary strand synthesis following rolling circle replication (Hafner *et al.*, 1997a). The 5' region of the CR-M, which does not contain the conserved GC-rich sequence, appears to be the binding site of an oligonuclotide primer that is involved in secondary strand synthesis in BBTV (Hafner *et al.*, 1997a) and likely the other two babuvirus species too.

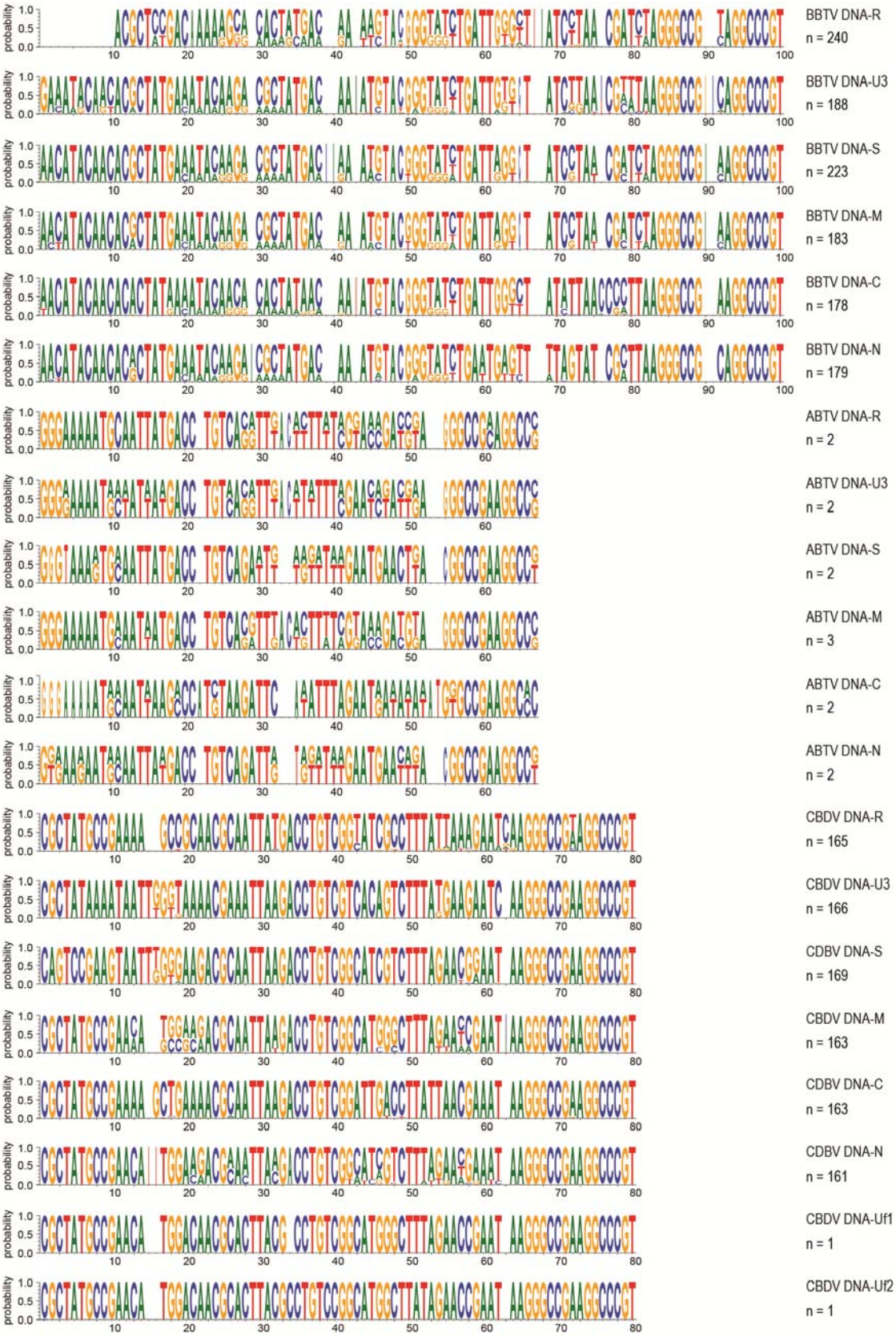


Figure 4.2: CR-M sequences of canonical babuvirus genome components. Each species was aligned separately. The widths of the letters take into consideration the number of gaps at that site across all the isolates in the alignment, for instance when there are a large number of gaps at a specific site relative to the number of nucleotides, the width of the letter is narrow.

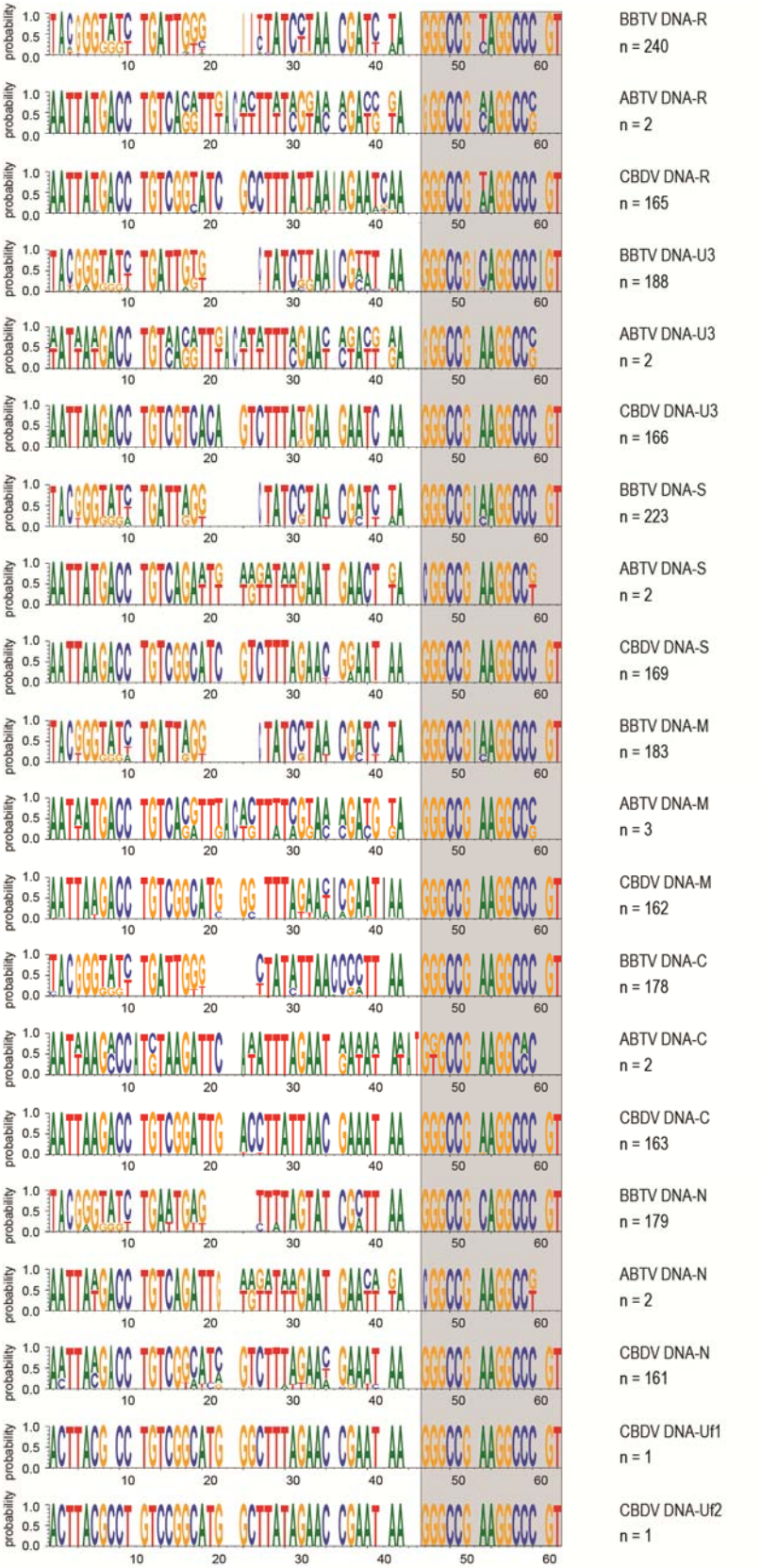


Figure 4.3: The 3' section of the CR-M of canonical babuvirus components which is conserved across the genome components. All species were aligned together. The GC-rich region is shown in grey. The widths of the letters take into consideration the number of gaps at that site across all the isolates in the alignment, for instance when there are a large number of gaps at a specific site relative to the number of nucleotides, the width of the letter is narrow.

4.4.3 Alphasatellite common regions

BBTV and CBDV have been found associated with alphasatellites. We therefore attempted to identify the CR-SL and CR-M sequences of all available geminivirus and nanovirus associated alphasatellites so that we could compare these to the DNA-R CR-SL and CR-M sequences of BBTV, CBDV and ABTV (Figure 4.4, Figure 4.5). We were, however, unable to identify the CR-SL sequences of alphasatellites that are associated with geminiviruses. The DNA-R components of the babuviruses all have the consensus (and highly conserved) *v-ori* nonanucleotide TATTATTAC (which is also found across all other babuvirus genome components; Figure 4.1), whereas the consensus *v-ori* nonanucleotide of the available alphasatellites is TAGTATTAC (Figure 4.4). Interestingly alphasatellites that are associated with babuviruses have the same nonanucleotide sequence as the other alphasatellites. These are also more similar across their entire CR-SL sequences to the other alphasatellites than they are to the CR-SL of babuvirus DNA-R sequences. This suggests that the babuvirus associated alphasatellites are likely not evolutionarily derived from babuvirus DNA-R sequences. The coconut foliar decay alphasatellite CR-SL is more similar to the alphasatellites than it is to the CR-SL sequences of babuvirus DNA-R component further supporting the hypothesis that CFDA is an alphasatellite with a currently undetermined associated virus rather than the DNA-R of a nanovirus as was originally assumed (King *et al.*, 2011; Rohde *et al.*, 1990).

Unlike the babuvirus DNA-R sequences, no iteron sequences homologous to iterons R, F1 and F2 could be unambiguously discerned within the CR-SLs of the various babuvirus-associated alphasatellites. Although 5 nt long sequences resembling these iterons were found 5' of the alphasatellite CR-SL sequences in some sequences (data not shown), these were not obviously conserved either within or between the various different groups of alphasatellites and were located further from the CR-SL than would be expected for functional iterons. These were not statistically common than would be expected for any random 5 nt long sequence (expected frequency = one in every 1024 nt vs observed frequency = one in every 1292 nt).

The CR-M sequences of the alphasatellites, like those of the babuviruses, displayed very little conservation between the different groups outside of a specific GC-rich region in the 5' or 3' portions of the CR-M (Figure 4.5). Two distinct groups of BBTV alphasatellite CR-M sequences were identified, with group-2 sequences (which all share >80% pairwise identity)

being more similar to cardamom bushy dwarf alphasatellite CR-M sequences (>71% pairwise identity) than they were to those of group-1 sequences (>67% pairwise identity; Figure 4.5).

Two distinct groups of geminivirus associated alphasatellites were also evident. Group-2 contains most of the available geminivirus-associated alphasatellite sequences (155 sequences) with a GC rich region at the 3' end of the CR-M. In contrast, group-1 contains a much shorter sequence before the GC-rich region, with a more conserved region after the GC rich region (Figure 4.5). Although some similarity is observable in the sequence preceding the GC-rich regions of the group-1 and group-2 geminivirus-associated alphasatellites, there is also a large amount of diversity, both across and within the groups, in that region. Interestingly, the 5' region of group-2 sequences is very similar to the 3' region of the group-1 sequences (underlined in black in Figure 4.5), with both regions being relatively conserved within each group, suggesting that the conservation of this sequence is important in all geminivirus associated alphasatellites. This adenine rich (A-rich) region has been identified previously in both geminivirus associated alphasatellites and geminivirus associated betasatellites and is thought to be involved either in complementary strand synthesis (Briddon *et al.*, 2003), or in increasing the genome size to approximately half that of a full size geminivirus genome component so as to ensure efficient encapsidation (Mansoor *et al.*, 1999; Saunders & Stanley, 1999). As the 5' end of the CR-M in BBTV has been identified as a primer binding site for complementary strand synthesis (Hafner *et al.*, 1997a), it is plausible that this conserved A-rich region might also act as a primer binding site in geminivirus-associated alphasatellites. However, the underlying reason for this conserved region being located in different locations within these alphasatellites in relation to the conserved GC-rich region remains unclear.

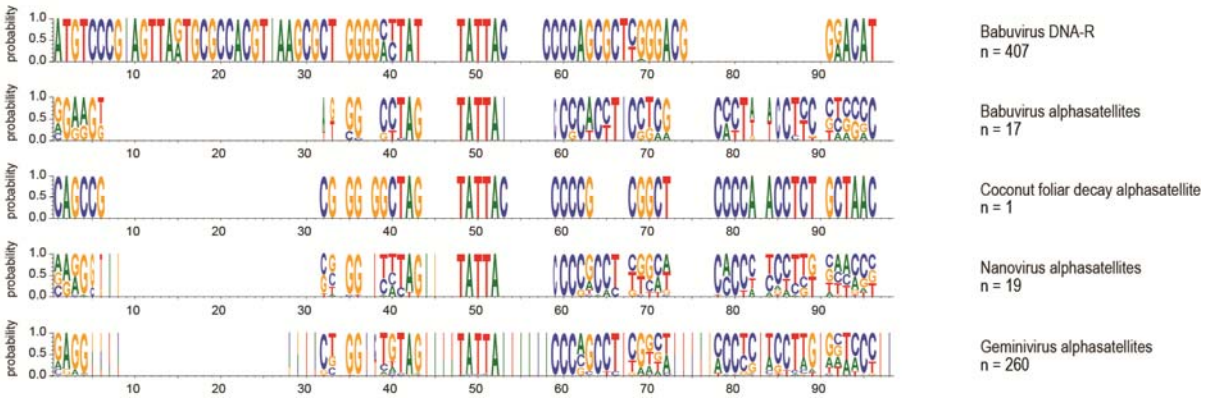


Figure 4.4: The CR-SL of the alphasatellites along with the CR-SL sequences of babuvirus DNA-R components for comparison. The widths of the letters take into consideration the number of gaps at that site across all the isolates in the alignment, for instance when there are a large number of gaps at a specific site relative to the number of nucleotides, the width of the letter is narrow.

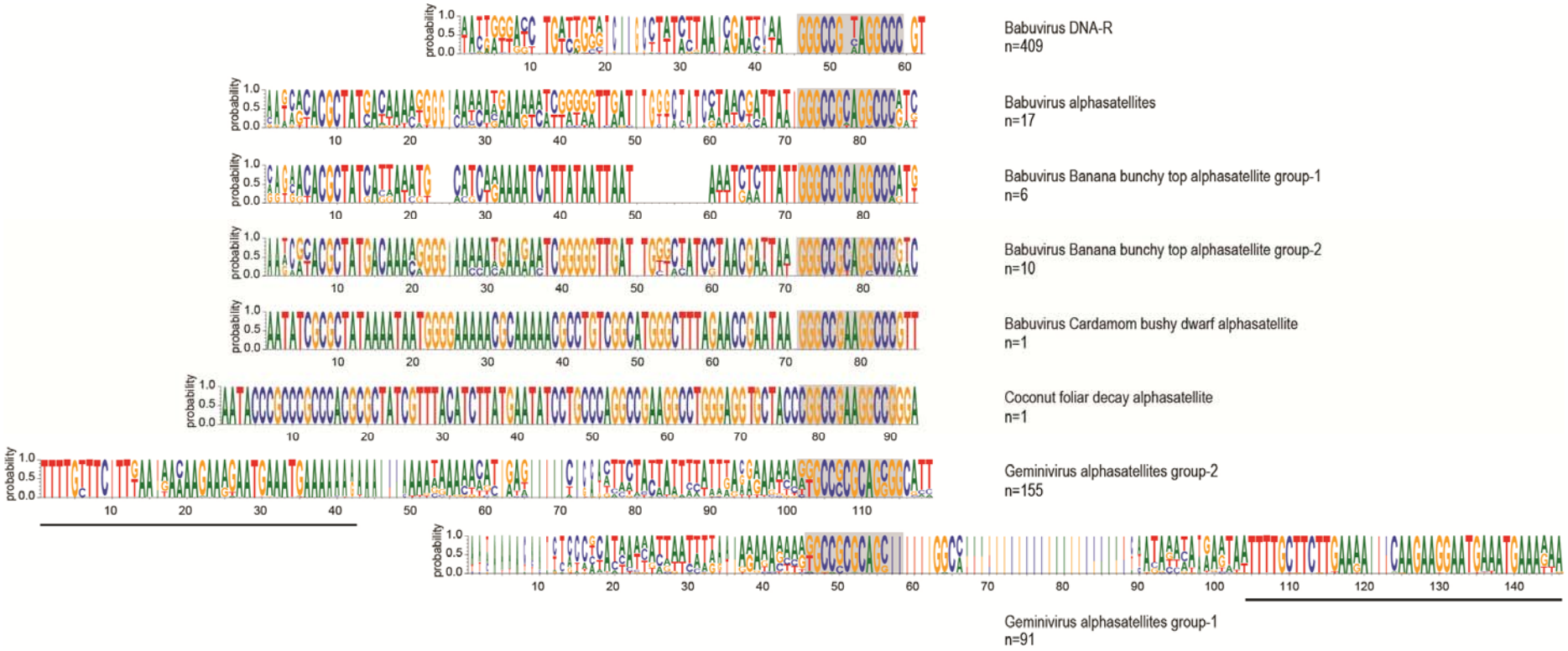


Figure 4.5: The CR-M of the alphasatellites with the CR-M of babuvirus DNA-R components for comparison. BBTV and geminivirus-associated alphasatellites have been split into two groups based on the degrees of similarity between their CR-M sequences. The black horizontal line indicates the region of sequence conservation between the geminivirus-associated alphasatelite groups. The GC-rich region is shown in grey. The widths of the letters take into consideration the number of gaps at that site across all the isolates in the alignment, for instance when there are a large number of gaps at a specific site relative to the number of nucleotides, the width of the letter is narrow.

4.4.4 Defective molecules

Twenty four defective BBTV genome components were serendipitously identified during efforts to clone and sequence the full genome sequences of BBTV isolates from various parts of the world (Stainton *et al.*, 2012; Stainton *et al.*, 2015). These defective molecules were amplified by PCR using BBTV-specific forward and reverse primers. The presence of binding sites for these primers in the defective BBTV genome components was a prerequisite for their cloning implying that the recovered molecules were not an unbiased sample of defective genome components that are present within BBTV infections. These defective genome components contained a range of deletions and/or insertions (Figure 4.6). Of the 24 analysed molecules 23 were determined to be defective DNA-R components, and one a defective DNA-U3 component. These molecules ranged in size from 544 nt to 1076 nt and were identified in banana leaf samples from seven different countries: Australia, Burundi, Philippines, Tonga, Taiwan, Sri Lanka, and Western Samoa. Isolate information, including accession numbers of associated BBTV isolates can be found in Table 4.2. Interestingly, we were unable to retrieve any other BBTV components from the plant samples from which d482P1-AU and dIP3-AU were isolated.

A common feature of the defective DNA-R molecules is that all, except dQ280-WS, have a deletion within the Rep ORF. The defective dQ280 DNA-R contains an intact Rep and all associated motifs, as well as a potential rep gene TATA box, suggesting that this ORF may be transcribed. A deletion in dQ280 has resulted in this molecule missing the 3' end of the CR-M and the 5' end of the CR-SL including the *v-ori* hairpin sequence. However, dQ280 does contain the iterons F1 (GGGAC) and F2 (GGGAC). Hence it is unknown whether this defective molecule is able to replicate.

All other defective DNA-R molecules contain a CR-M and CR-SL which are similar to those found in non-defective BBTV genome components and are therefore likely to be trans-replicated by the Rep of the cognate BBTV. The only non DNA-R defective molecule, dKP5 contains an intact CR-M and has an insert derived from a BBTV alphasatellite within the CR-SL, which has resulted in this molecule having an alphasatellite-like CR-SL. Therefore, although this molecule is unlikely to be replicated by BBTV DNA-R, it is plausible that the Rep of an alphasatellite may recognise and replicate dKP5.

Similar deletions and insertions were observed in multiple different isolates sampled from different plants. For example, two molecules from Australia d1429A and d482P1 have two

identical deletions and 11 Tongan sequences (dTOS55, dTOS58, dTOS59, dTOS61, dTOS64, dTOS67, dTOS69, dTOS80, dTOS82, dTOS87, dTOS92) have similar patterns of deletion and insertion. Seven of these 11 isolates (dTOS55, dTOS58, dTOS61, dTOS64, dTOS67, dTOS82, dTOS92) also possess an insertion consisting of a repeating sequence 12-13 nt long. Two breakpoint (deletion or insertion) hotspots were identified in the DNA-R defective molecules (Figure 4.7).

A ML phylogenetic tree was constructed with all recombination free DNA-R sequences and defective molecules with insert sequences removed, to elucidate possible origins of the defective molecules. Defective molecules grouped with DNA-R sequences from the same country (Figure 4.8). Two clades, both with high statistical support, contained only defective molecules. One of the clades contained the seven isolates mentioned above dTOS55, dTOS58, dTOS61, dTOS64, dTOS67, dTOS82 and dTOS92. The other contained dMPI, d626M, d626 and dMP2, suggesting a common lineage for the isolates of each of the two clades. The identification of these common lineages demonstrates that defective molecules are being replicated and moved between hosts. Further evidence for the movement of these molecules is the isolate dMP2 and the DNA-R isolate MP2, which were identified from the same plant, but are not found on the same lineage.

Other defective molecules were seen with insertions consisting of repeating sequences, these sequences were either direct repeats or the repeat unit occurred further upstream or downstream. Repeat regions are possibly due to slippage events of the products of DNA polymerase complex during replication. Six insertions in the defective molecules (dTOS59, dTOS61, dTOS78, dTOS87, dIP3, dKP5) appear to have been derived from a different component. Besides the BBTV alphasatellite insertion in dKP5 DNA-U3 mentioned above, we observed two other DNA-U3 insertions, two DNA-C insertions and one DNA-N insertion within the defective DNA-R molecules. These defective molecules have likely arisen through recombination which suggests that the defective molecules may serve as a further source of genetic material for BBTV.

As all defective molecules were opportunistically identified, defective molecules could potentially be a common occurrence in BBTV infections. A number of defective geminivirus molecules have been identified, as reviewed in Patil and Dasgupta (2006) and Rey *et al.* (2012), which reduce or delay the viral symptoms in infected host plants, which is thought to

be due to increased competition for resources such as the DNA binding sites of Rep proteins (Idris *et al.*, 2011; Ndunguru *et al.*, 2006; Stanley *et al.*, 1990).

However, in some cases at least, the presence of defective geminivirus molecules within an infected plant does not cause any reduced symptoms (Schubert *et al.*, 2014). It is currently unknown whether BBTV defective molecules have an influence on the severity of BBTV infections.

Table 4.2: Isolate information for defective genomes. Accession numbers of the associated integral components of BBTV are given where identified. Two letter country codes are given, AU - Australia, BI - Burundi, TW - Taiwan, TO - Tonga, LK - Sri Lanka, WS - Samoa.

Isolate	Country	Year	Defective molecule	DNA-R Accession	DNA-U3 Accession	DNA-S Accession	DNA-M Accession	DNA-C accession	DNA-N Accession
1429A	AU		d1429A			KM607439	KM607148	KM607005	KM607294
482P1	AU	2011	d482P1						
547	BI	1995	d547		KM607735	KM607454	KM607164	KM607020	KM607309
KP5	LK	2003	dKP5	KM607656		KM607512	KM607218	KM607076	KM607364
MS14	PH	2008	dMS14		KM607790	KM607517	KM607224	KM607082	
TOS55	TO	2010	dTOS55		KM607831	KM607560	KM607265	KM607122	KM607411
TOS58	TO	2010	dTOS58		KM607833	KM607562	KM607268	KM607125	KM607414
TOS59	TO	2010	dTOS59		KM607834	KM607563	KM607269	KM607126	
TOS61	TO	2010	dTOS61		KM607836	KM607565		KM607128	KM607416
TOS64	TO	2010	dTOS64		KM607840	KM607569	KM607274	KM607132	KM607419
TOS67	TO	2010	dTOS67		KM607842		KM607276	KM607134	KM607421
TOS69	TO	2010	dTOS69		KM607844		KM607278	KM607136	
TOS71	TO	2010	dTOS71		KM607845	KM607573	KM607280	KM607138	KM607424
TOS78	TO	2010	dTOS78		KM607848	KM607574	KM607283	KM607141	KM607427
TOS80	TO	2010	dTOS80		KM607850		KM607284		KM607428
TOS82	TO	2010	dTOS82		KM607851	KM607575	KM607285	KM607142	KM607429
TOS87	TO	2010	dTOS87		KM607854	KM607578	KM607288	KM607144	KM607433
TOS92	TO	2010	dTOS92			KM607583			
dIP3	TW	1996	dIP3						
626	TW	1996	d626			KM607463			KM607318
626M	TW	1995	d626M			KM607464			
MP1	TW	1996	dMP1		KM607788	KM607515			KM607368
MP2	TW	1996	dMP2	KM607661	KM607789	KM607516	KM607223	KM607081	KM607369
Q280	WS	1989	dQ280						KM607381

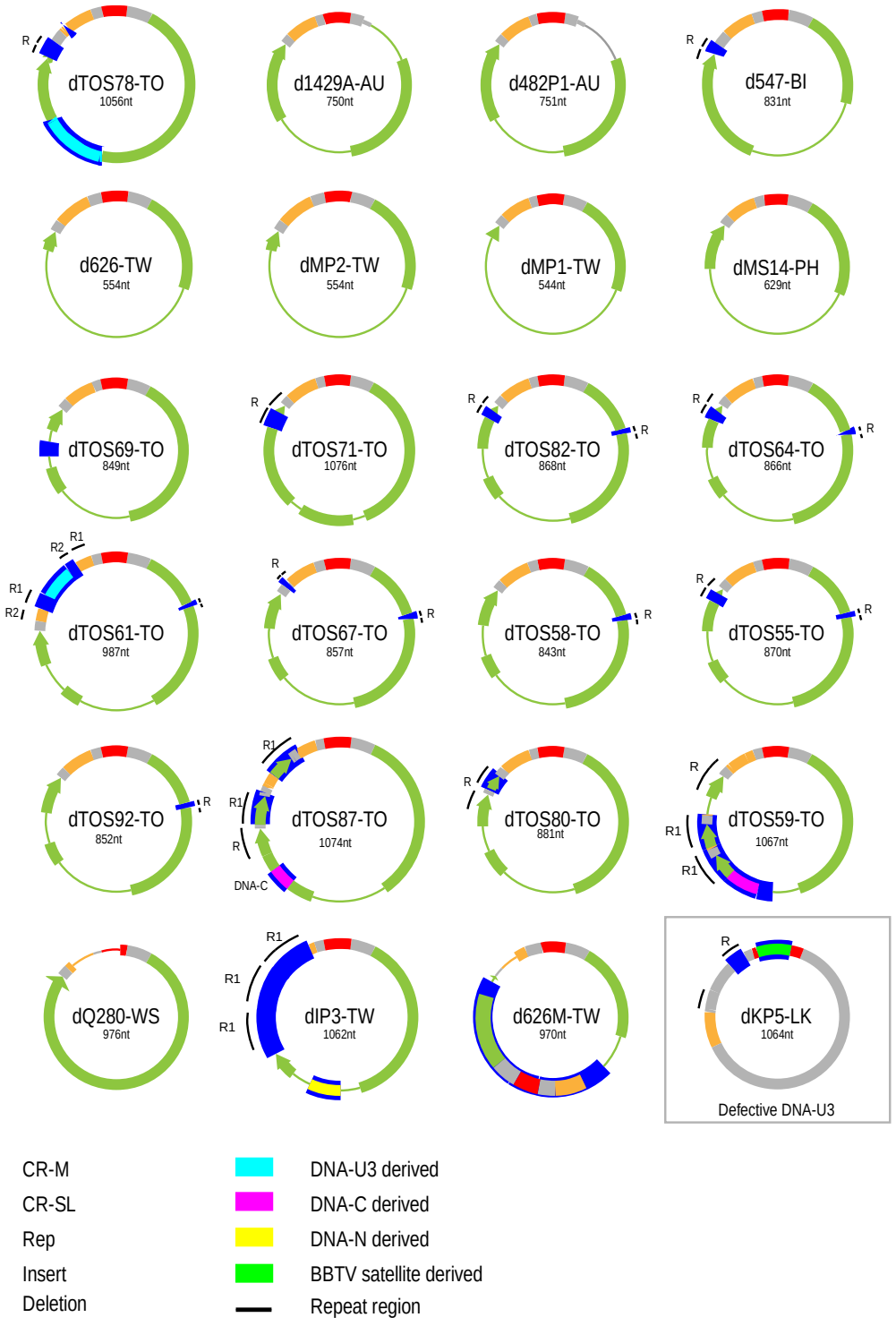


Figure 4.6: Illustration of defective molecules in relation to the reference sequence, TOS60-R (KM607707), or, in the case of KP5-U3, to TOS60-U3 (KM607835). The size is relative to the reference sequences with the exact size in nucleotides stated for each molecule. Two letter country codes are shown wherever these are known, AU - Australia, BI - Burundi, TW - Taiwan, TO - Tonga, LK - Sri Lanka, WS - Samoa. The CR-SL and common region major CR-M, relative to the reference sequence are shown in red and orange respectively, with the rest of the non-coding region shown in grey. The replication associated protein is shown in green. Deletions are shown as thin lines with insertions shown in blue. Repeat sequences are shown with the letter R and inserts which contain sequences derived from other genome components or alphasatellites are shown in specific colours.

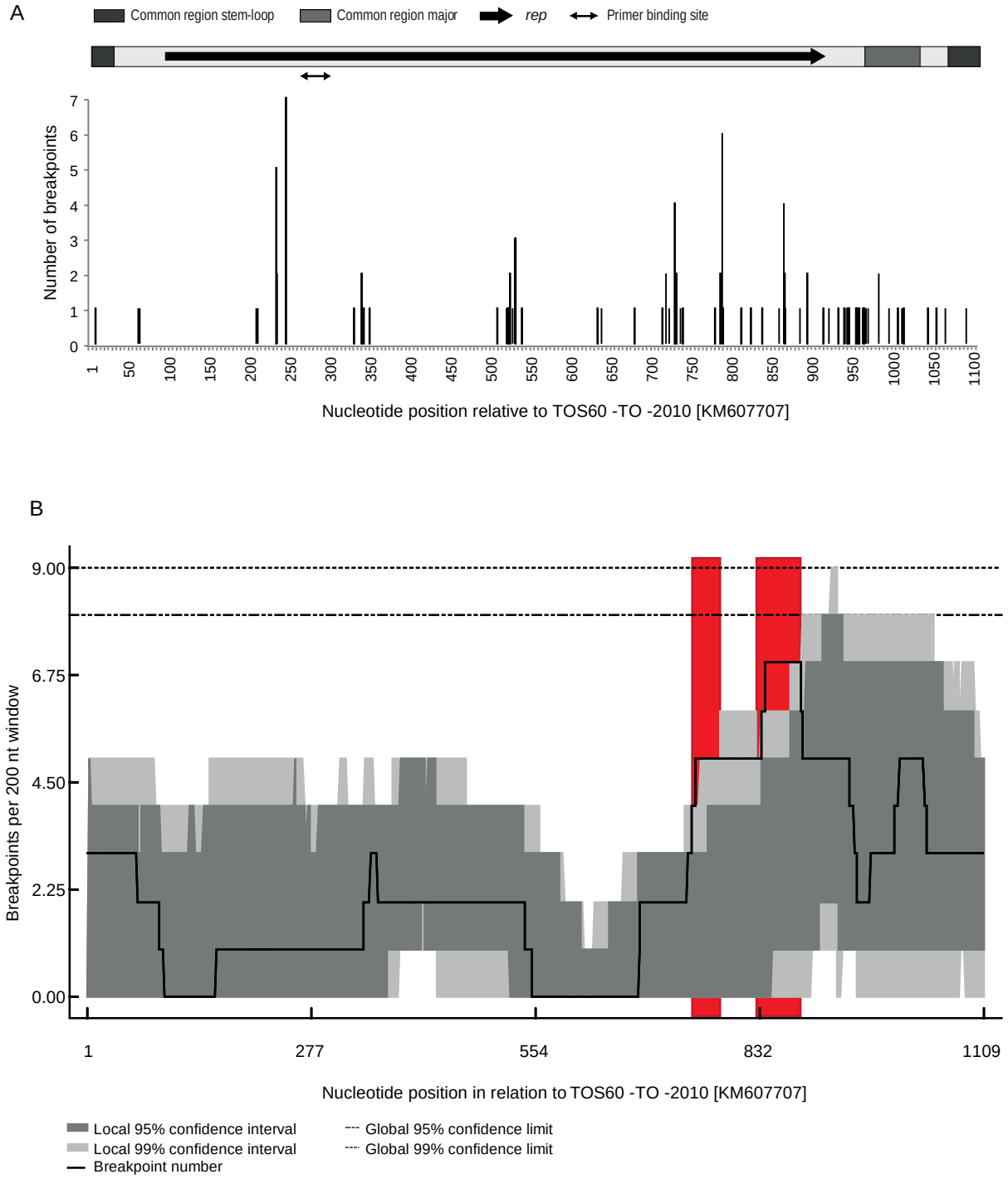


Figure 4.7: A) Individual breakpoint positions relative to the DNA-R component architecture. The back-to-back primer binding site is shown. B) Recombination breakpoint distribution plot of DNA-R with 95% and 99% confidence intervals. Solid black lines show identified breakpoints. Two breakpoint hotspots are identified (red) where breakpoint numbers are above local 99% confidence levels.

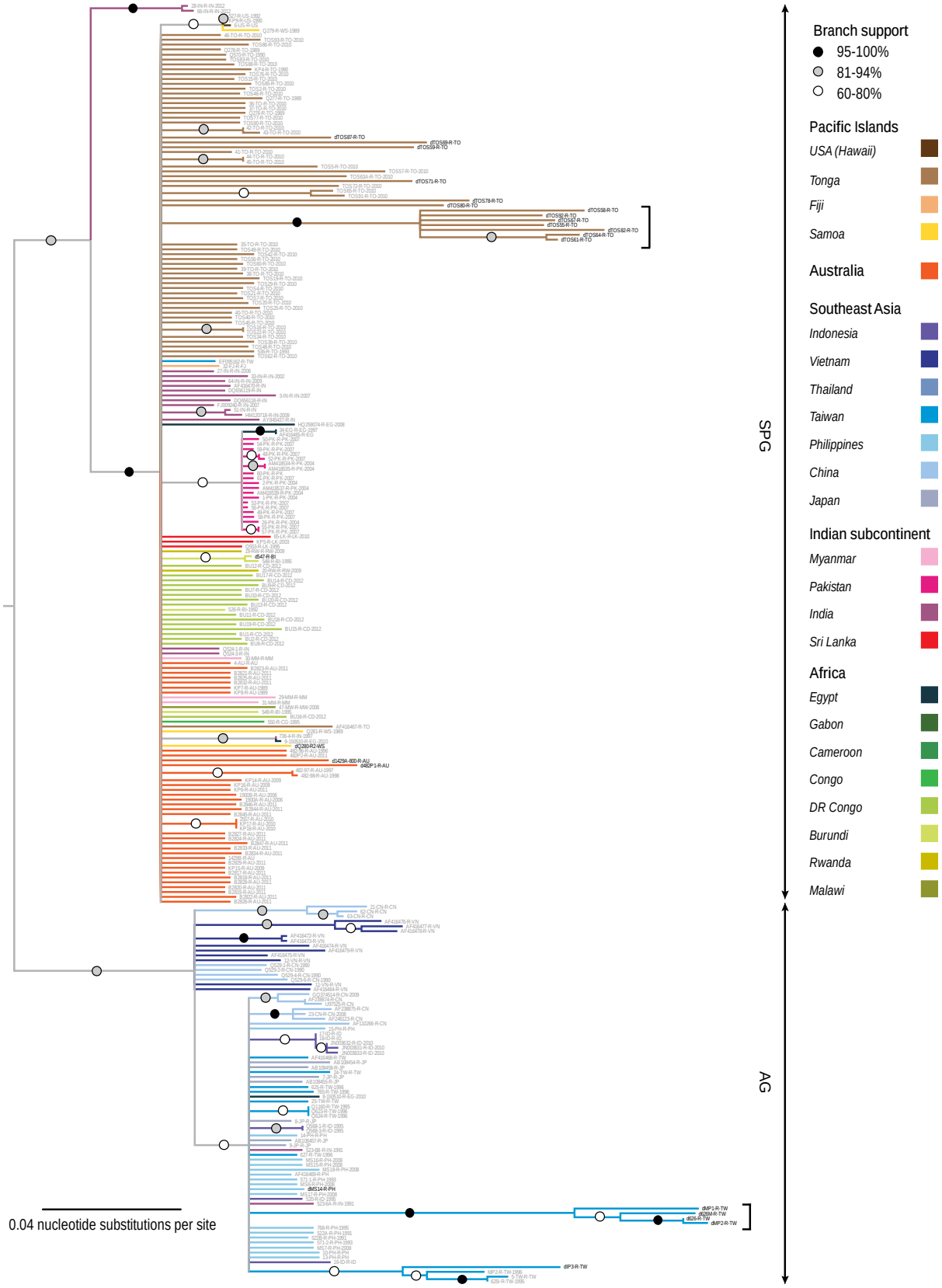


Figure 4.8: Maximum likelihood phylogenetic tree (inferred using model T92+G) of defective DNA-R molecules (insert regions removed) and DNA-R components (recombination free). The phylogenetic tree was rooted with ABTV DNA-R sequences and branches with < 60% bootstrap supports collapsed. Isolate information is available in Chapter Three (Supplementary Table 3.1) for the DNA-R sequences and Table 4.1 for the defective genomes.

4.5 Concluding remarks

Full CR-SL sequences were identified from all canonical babuvirus genome components. The CR-SL of the canonical babuvirus genome components were similar across all three species, suggesting that the DNA-R components may be able to trans-replicate the genome components of all three species. The CR-M, however, was not highly conserved across all the analysed babuviruses species with the highest degree of similarity between the different species being detectable in a GC-rich at the 3' end region of the CR-M.

The CR-SL sequences of the geminivirus-, nanovirus- and babuvirus-associated alphasatellite molecules are more similar to one another than any are to the babuvirus DNA-R component, suggesting that the Reps expressed from these molecules are probably incapable of trans-replicating one another. Our analyses also confirms that, based on the CR-SL sequence, the coconut foliar decay component appears to be an alphasatellite rather than the DNA-R of an uncharacterised nanovirus, however the associated virus is yet to be identified. Like the babuvirus components, the alphasatellite CR-M was only conserved at the GC rich region.

Defective BBTV-associated DNA-R molecules were characterised to reveal that the vast majority of disruptions in this component occur in the Rep encoding sequences, with intact CR-SL and CR-M. However, as these were identified through PCR based sampling, the primers may result in a sampling bias. The identifiable insert sequences were either derived from other components, or as the result of repeating sequences, which suggests recombination or slippage events respectively occurred during replication. Recombination was also seen between a defective molecule and an alphasatellite indicating both that the alphasatellites and defective molecules are present during recombination and that they are potential sources of genetic diversity for continuing diversification of BBTV. Two breakpoint hotspots were identified, and the ML phylogenetic tree suggests a number of defective sequences share a common lineage.

Supplementary Table 4.1: Accession numbers of the sequences used in the CR-M and CR-SL datasets for the canonical BBTV, ABTV and CBDV isolates. Except for a small number of CBDV sequences, both the CR-M and CR-SL were analysed from all isolates.

BBTV						ABTV							
DNA-R	DNA-U3	DNA-S	DNA-M	DNA-C	DNA-N	DNA-R	DNA-U3	DNA-S	DNA-M	DNA-C	DNA-N		
		KM607467	KM607173	KM607030	KM607321	EF546813	EF546809	EF546810	EF546811	EF546812	EF546808		
KM607588	KM607724	KM607442	KM607152	KM607009	KM607298	EF546807	EF546803	EF546804	EF546805	EF546806	EF546802		
		KM607439	KM607148	KM607005	KM607294				AF102148				
KM607585			KM607149	KM607006	KM607295								
KM607586	KM607722	KM607440	KM607150	KM607007	KM607296	CBDV							
KM607587	KM607723	KM607441	KM607151	KM607008	KM607297	DNA-R	DNA-U3	DNA-S	DNA-M	DNA-C	DNA-N	DNA-Uf1	DNA-Uf2
KM607589	KM607725	KM607443	KM607153	KM607010	KM607299	KF710463	KF710626	KF710789	KF10952	KF711115	KF711278		
KM607590	KM607726	KM607444	KM607154	KM607011	KM607300	KF710464	KF710627	KF710790	KF10953	KF711116	KF711279		
KM607591	KM607727	KM607445	KM607155	KM607012	KM607301	KF710465	KF710628	KF710791	KF10954	KF711117	KF711280		
KM607592						KF710466	KF710629	KF710792	KF10955	KF711118	KF711281		
S56276	L41576	L41574	L41575	L41578	L41577	KF710467	KF710630	KF710793	KF10956	KF711119	KF711282		
KM607614	KM607749	KM607472	KM607178	KM607035	KM607325	KF710468	KF710631	KF710794	KF10957	KF711120			
KM607615	KM607750		KM607179	KM607036	KM607326	KF710469	KF710632	KF710795	KF10958	KF711121	KF711284		
KM607616	KM607751	KM607473	KM607180	KM607037	KM607327	KF710470	KF710633	KF710796	KF10959	KF711122	KF711285		
KM607617		KM607474	KM607181	KM607038		KF710471	KF710634	KF710797	KF10960	KF711123	KF711286		
KM607618	KM607752	KM607475	KM607182	KM607039	KM607328	KF710472	KF710635	KF710798	KF10961	KF711124	KF711287		
KM607619	KM607753	KM607476	KM607183	KM607040	KM607329	KF710473	KF710636	KF710799	KF10962	KF711125	KF711288		
KM607620	KM607754	KM607477		KM607041	KM607330	KF710474	KF710637	KF710800	KF10963	KF711126	KF711289		
KM607621	KM607755	KM607478	KM607184	KM607042	KM607331	KF710475	KF710638	KF710801	KF10964	KF711127	KF711290		
KM607622	KM607756	KM607479	KM607185	KM607043	KM607332	KF710476	KF710639	KF710802	KF10965	KF711128	KF711291		
KM607623		KM607480	KM607186		KM607333	KF710477	KF710640	KF710803	KF10966	KF711129	KF711292		
KM607624	KM607757	KM607481		KM607044		KF710478	KF710641	KF710804	KF10967	KF711130	KF711293		
KM607625		KM607482	KM607187	KM607045		KF710479	KF710642	KF710805	KF10968	KF711131	KF711294		
KM607626	KM607758	KM607483	KM607188	KM607046	KM607334	KF710480	KF710643	KF710806	KF10969	KF711132	KF711295		
	KM607759	KM607484	KM607189	KM607047	KM607335	KF710481	KF710644	KF710807	KF10970	KF711133	KF711296		
KM607627		KM607485	KM607190	KM607048		KF710482	KF710645	KF710808	KF10971	KF711134	KF711297		
KM607628	KM607760		KM607191	KM607049	KM607336	KF710483	KF710646	KF710809	KF10972	KF711135	KF711298		
KM607629	KM607761	KM607486	KM607192		KM607337	KF710484	KF710647	KF710810	KF10973	KF711136	KF711299		
KM607630				KM607050	KM607338	KF710485	KF710648	KF710811	KF10974	KF711137	KF711300		
KM607631	KM607762	KM607487	KM607193	KM607051	KM607339	KF710486	KF710649	KF710812	KF10975	KF711138	KF711301		
KM607632		KM607488	KM607194	KM607052	KM607340	KF710487	KF710650	KF710813	KF10976	KF711139	KF711302		
KM607633	KM607763	KM607489	KM607195	KM607053	KM607341	KF710488	KF710651	KF710814	KF10977 NO CRM	KF711140	KF711303		
KM607650	KM607780	KM607506	KM607212	KM607070	KM607358	KF710489	KF710652	KF710815	KF10978	KF711141	KF711304		
KM607651	KM607781	KM607507	KM607213	KM607071	KM607359	KF710490	KF710653	KF710816	KF10979	KF711142	KF711305		
KM607653	KM607783	KM607509	KM607215	KM607073	KM607361	KF710491	KF710654	KF710817	KF10980	KF711143	KF711306		
KM607654	KM607784	KM607510	KM607216	KM607074	KM607362	KF710492	KF710655	KF710818	KF10981	KF711144	KF711307		
KM607657	KM607786	KM607513	KM607219	KM607077	KM607365	KF710493	KF710656	KF710819	KF10982	KF711145	KF711308		
KM607658			KM607220	KM607078		KF710494	KF710657	KF710820	KF10983	KF711146	KF711309		
KM607659			KM607221	KM607079	KM607366	KF710495	KF710658	KF710821	KF10984	KF711147	KF711310		
	KM607741	KM607460	KM607170	KM607026	KM607315	KF710496	KF710659	KF710822	KF10985	KF711148	KF711311		
KM607652	KM607782	KM607508	KM607214	KM607072	KM607360	KF710497	KF710660	KF710823	KF10986	KF711149	KF711312		
KM607598	KM607732	KM607451	KM607161	KM607017	KM607306	KF710498	KF710661	KF710824	KF10987	KF711150	KF711313		
	KM607735	KM607454	KM607164	KM607020	KM607309	KF710499	KF710662	KF710825	KF10988	KF711151	KF711314		
KM607601	KM607736	KM607455	KM607165	KM607021	KM607310	KF710500	KF710663	KF710826	KF10989	KF711152	KF711315		
KM607602	KM607737	KM607456	KM607166	KM607022	KM607311	KF710501	KF710664	KF710827	KF10990	KF711153	KF711316		
		AF148943				KF710502	KF710665	KF710828	KF10991	KF711154	KF711317		
KM607634	KM607764	KM607490	KM607196	KM607054	KM607342	KF710503	KF710666	KF710829	KF10992	KF711155			
KM607635	KM607765	KM607491	KM607197	KM607055	KM607343	KF710504	KF710667	KF710830	KF10993	KF711156	KF711319		
KM607636	KM607766	KM607492	KM607198	KM607056	KM607344	KF710505	KF710668	KF710831	KF10994	KF711157	KF711320		
KM607637	KM607767	KM607493	KM607199	KM607057	KM607345	KF710506	KF710669	KF710832	KF10995	KF711158	KF711321		
KM607638	KM607768	KM607494	KM607200	KM607058	KM607346	KF710507	KF710670	KF710833	KF10996	KF711159	KF711322		

KM607639 KM607769 KM607495 KM607201 KM607059 KM607347
KM607640 KM607770 KM607496 KM607202 KM607060 KM607348
KM607641 KM607771 KM607497 KM607203 KM607061 KM607349
KM607642 KM607772 KM607498 KM607204 KM607062 KM607350
KM607643 KM607773 KM607499 KM607205 KM607063 KM607351
KM607644 KM607774 KM607500 KM607206 KM607064 KM607352
KM607645 KM607775 KM607501 KM607207 KM607065 KM607353
KM607646 KM607776 KM607502 KM607208 KM607066 KM607354
KM607647 KM607777 KM607503 KM607209 KM607067 KM607355
KM607648 KM607778 KM607504 KM607210 KM607068 KM607356
KM607649 KM607779 KM607505 KM607211 KM607069 KM607357
JF755984
JF755986
JF755987
KM607603 KM607738 KM607457 KM607167 KM607023 KM607312
GQ249344
JF755978
AY450396 AY606084 AY494786 AY494788 AY606085 AY494787
FJ463042 FJ463043 FJ463044 FJ463045 FJ463046 FJ463047
HQ378190 HM231314 HQ378191 HQ378192 HQ378193 HQ378194
HQ616074 HQ616075 HQ616076 HQ616077 HQ616078 HQ616079
AF110266
AF238874
AF238875
AF238876
AF238877
AF238878
AF238879
AF246123
AF349568
AY264347
AY266417
EF470243
GQ374514
GU559702
GU559703
GU559704
GU559705
GU559706
HM212635
KM607676 KM607533 KM607385
KM607677 KM607806 KM607534 KM607098 KM607386
KM607678 KM607807 KM607535 KM607239 KM607099 KM607387
KM607536
KM607679 KM607808 KM607240
U97525
AF102780 AF102781 AF102783
KM607612 KM607747 KM607470 KM607176 KM607033 KM607324
KM607613 KM607748 KM607471 KM607177 KM607034
AF102784
AF416465
HQ259074
AF416466 AF148944
JF755981
JF755982
KM607593 KM607728 KM607446 KM607156 KM607013 KM607302
AB186924 AB186927
AB186925 AB186928

KF710508 KF710671 KF710834 KF10997 KF711160 KF711323
KF710509 KF710672 KF710835 KF10998 KF711161 KF711324
KF710510 KF710673 KF710836 KF10999 KF711162 KF711325
KF710511 KF710674 KF710837 KF11000 KF711163 KF711326
KF710512 KF710675 KF710838 KF11001 KF711164 KF711327
KF710513 KF710676 KF710839 KF11002 KF711165 KF711328
KF710514 KF710677 KF710840 KF11003 KF711166
KF710515 KF710678 KF710841 KF11004 KF711167 KF711330
KF710516 KF710679 KF710842 KF11005 KF711168 KF711331
KF710517 KF710680 KF710843 KF11006 KF711169 KF711332
KF710518 KF710681 KF710844 KF11007 KF711170 KF711333
KF710519 KF710682 KF710845 KF11008 KF711171 KF711334
KF710520 KF710683 KF710846 KF11009 KF711172 KF711335
KF710521 KF710684 KF710847 KF11010 KF711173 KF711336
KF710522 KF710685 KF710848 KF11011 KF711174 KF711337
KF710523 KF710686 KF710849 KF11012 KF711175 KF711338
KF710524 KF710687 KF710850 KF11013 KF711176 KF711339
KF710525 KF710688 KF710851 KF11014 KF711177 KF711340
KF710526 KF710689 KF710852 KF11015 KF711178 KF711341
KF710527 KF710690 KF710853 KF11016 KF711179 KF711342
KF710528 KF710691 KF710854 KF11017 KF711180 KF711343
KF710529 KF710692 KF710855 KF11018 KF711181 KF711344
KF710530 KF710693 KF710856 KF11019 KF711182 KF711345
KF710531 KF710694 KF710857 KF11020 KF711183 KF711346
KF710532 KF710695 KF710858 KF11021 KF711184 KF711347
KF710533 KF710696 KF710859 KF11022 KF711185 KF711348
KF710534 KF710697 KF710860 KF11023 KF711186 KF711349
KF710535 KF710698 KF710861 KF11024 KF711187 KF711350
KF710536 KF710699 KF710862 KF11025 KF711188 KF711351
KF710537 KF710700 KF710863 KF11026 KF711189 KF711352
KF710538 KF710701 KF710864 KF11027 KF711190 KF711353
KF710539 KF710702 KF710865 KF11028 KF711191 KF711354
KF710540 KF710703 KF710866 KF11029 KF711192 KF711355
KF710541 KF710704 KF710867 KF11030 KF711193 KF711356
KF710542 KF710705 KF710868 KF11031 KF711194 KF711357
KF710543 KF710706 KF710869 KF11032 KF711195 KF711358
KF710544 KF710707 KF710870 KF11033 KF711196 KF711359
KF710545 KF710708 KF710871 KF11034 KF711197 KF711360
KF710546 KF710709 KF710872 KF11035 KF711198 KF711361
KF710547 KF710710 KF710873 KF11036 KF711199 KF711362
KF710548 KF710711 KF710874 KF11037 KF711200 KF711363
KF710549 KF710712 KF710875 KF11038 KF711201 KF711364
KF710550 KF710713 KF710876 KF11039 KF711202 KF711365
KF710551 KF710714 KF710877 KF11040 KF711203 KF711366
KF710552 KF710715 KF710878 KF11041 KF711367
KF710553 KF710716 KF710879 KF11042 KF711205 KF711368
KF710554 KF710717 KF710880 KF11043 KF711206 KF711369
KF710555 KF710718 KF710881 KF11044 KF711207 KF711370
KF710556 KF710719 KF710882 KF11045 KF711208 KF711371
KF710557 KF710720 KF710883 KF11046 KF711209 KF711372
KF710558 KF710721 KF710884 KF11047 KF711210 KF711373
KF710559 KF710722 KF710885 KF11048 KF711211 KF711374
KF710560 KF710723 KF710886 KF11049 KF711212 KF711375
KF710561 KF710724 KF710887 KF11050 KF711213 KF711376
KF710562 KF710725 KF710888 KF11051 KF711214 KF711377
KF710563 KF710726 KF710889 KF11052 KF711215 KF711378
NO CRM
KF710564 KF710727 KF710890 KF11053 KF711216 KF711379
KF710565 KF710728 KF710891 KF11054 KF711217 KF711380

AB186926		AB186929			
JN003631					
JN003632					
JN003633					
KM607681	KM607810	KM607538	KM607242	KM607101	KM607389
KM607682	KM607811		KM607243	KM607102	KM607390
	AY960129		AY948439		AY948438
FJ009238	FJ009239				
JQ911667	JQ911668				
EU140342	EU140341	EU589459	EU190971	EU190969	EU190970
DQ256267	EU402601	EF687856	EU516323	EU051379	EU391633
FJ605506	FJ605508	FJ605507	FJ609642	FJ609643	FJ609644
KM607596		KM607449	KM607159		
KM607597	KM607731	KM607450	KM607160	KM607016	KM607305
JN243751	JN243752	JN243753			JN243754
KC119098	KC466373	KC466374	KC466375	KC466376	KC466377
KM607609	KM607745	KM607466	KM607172	KM607029	KM607320
AF416470					
AY845437					
	AY884172				
	AY884173				
			AY953429		
DQ656118					
DQ656119					
	EU046323				
FJ009240					
HM120718					
KM607674	KM607803	KM607530	KM607237	KM607095	KM607383
	KM607804	KM607531		KM607096	
KM607675	KM607805	KM607532	KM607238	KM607097	KM607384
AB108452		AB108449			
AB108453		AB108450			
AB108456		AB108451			
AB108454					
AB108455					
AB108457					
AB108458					
JN250593	JN250594	JN250595	JN250596	JN250597	JN250598
KM607656		KM607512	KM607218	KM607076	KM607364
KM607680	KM607809	KM607537	KM607241	KM607100	KM607388
AB252639		AB252642			
AB252640		AB252643			
AB252641		AB252644			
		JF755980			
JQ820453	JQ820454	JQ820455	JQ820456	JQ820457	JQ820458
KM607611		KM607469	KM607175	KM607032	KM607323
AB189067		AB189068			
AB250953		AB250956			
AB250954		AB250957			
AB250955		AB250958			
KM607594	KM607729	KM607447	KM607157	KM607014	KM607303
KM607595	KM607730	KM607448	KM607158	KM607015	KM607304
KM607604	KM607740	KM607459	KM607169	KM607025	KM607314
KM607605	KM607739	KM607458	KM607168	KM607024	KM607313
		AF148068			
AF416469					

KF710566	KF710729	KF710892	KF11055	KF711218	KF711381
KF710567	KF710730	KF710893	KF11056	KF711219	KF711382
KF710568	KF710731	KF710894	KF11057	KF711220	KF711383
KF710569	KF710732	KF710895	KF11058	KF711221	KF711384
KF710570	KF710733	KF710896	KF11059	KF711222	KF711385
KF710571	KF710734	KF710897	KF11060	KF711223	KF711386
KF710572	KF710735	KF710898	KF11061	KF711224	KF711387
KF710573	KF710736	KF710899	KF11062	KF711225	KF711388
KF710574	KF710737	KF710900	KF11063	KF711226	KF711389
KF710575	KF710738	KF710901	KF11064	KF711227	KF711390
KF710576	KF710739	KF710902	KF11065	KF711228	KF711391
KF710577	KF710740	KF710903	KF11066	KF711229	KF711392
KF710578	KF710741	KF710904	KF11067	KF711230	KF711393
KF710579	KF710742	KF710905	KF11068	KF711231	KF711394
KF710580	KF710743	KF710906	KF11069	KF711232	KF711395
KF710581	KF710744	KF710907	KF11070	KF711233	KF711396
KF710582	KF710745	KF710908	KF11071	KF711234	KF711397
KF710583	KF710746	KF710909	KF11072	KF711235	KF711398
KF710584	KF710747	KF710910	KF11073	KF711236	KF711399
KF710585	KF710748	KF710911	KF11074	KF711237	KF711400
KF710586	KF710749	KF710912	KF11075	KF711238	KF711401
KF710587	KF710750	KF710913	KF11076	KF711239	KF711402
KF710588	KF710751	KF710914	KF11077	KF711240	KF711403
KF710589	KF710752	KF710915	KF11078	KF711241	KF711404
KF710590	KF710753	KF710916	KF11079	KF711242	KF711405
KF710591	KF710754	KF710917	KF11080	KF711243	KF711406
KF710592	KF710755	KF710918	KF11081	KF711244	KF711407
KF710593	KF710756	KF710919	KF11082	KF711245	KF711408
KF710594	KF710757	KF710920	KF11083 NO CRM	KF711246	KF711409
KF710595	KF710758	KF710921	KF11084	KF711247	KF711410
KF710596	KF710759	KF710922	KF11085	KF711248	KF711411
KF710597	KF710760	KF710923	KF11086 NO CR-SL	KF711249	KF711412
KF710598	KF710761	KF710924	KF11087	KF711250	KF711413
KF710599	KF710762	KF710925	KF11088	KF711251	KF711414
KF710600	KF710763	KF710926	KF11089	KF711252	KF711415
KF710601	KF710764	KF710927	KF11090	KF711253	KF711416
KF710602	KF710765	KF710928	KF11091	KF711254	KF711417
KF710603	KF710766	KF710929	KF11092	KF711255	KF711418
KF710604	KF710767	KF710930	KF11093	KF711256	KF711419
KF710605	KF710768	KF710931	KF11094	KF711257	KF711420 NO CRM
KF710606	KF710769	KF710932	KF11095	KF711258	KF711421
KF710607	KF710770	KF710933	KF11096	KF711259	KF711422
KF710608	KF710771	KF710934	KF11097	KF711260	KF711423
KF710609	KF710772	KF710935	KF11098	KF711261	KF711424
KF710610	KF710773	KF710936	KF11099	KF711262	KF711425
KF710611	KF710774	KF710937	KF11100	KF711263	KF711426
KF710612	KF710775	KF710938	KF11101	KF711264	KF711427
KF710613	KF710776	KF710939	KF11102	KF711265	KF711428
KF710614	KF710777	KF710940	KF11103	KF711266	KF711429
KF710615	KF710778	KF710941	KF11104	KF711267	KF711430
KF710616	KF710779	KF710942	KF11105	KF711268	KF711431
KF710617	KF710780	KF710943	KF11106	KF711269	KF711432
KF710618	KF710781	KF710944	KF11107	KF711270	KF711433
KF710619	KF710782	KF710945	KF11108	KF711271	KF711434
KF710620	KF710783	KF710946	KF11109	KF711272	KF711435
KF710621	KF710784	KF710947	KF11110	KF711273	KF711436

	KM607790	KM607517	KM607224	KM607082	
KM607662	KM607791	KM607518	KM607225	KM607083	KM607370
KM607663	KM607792	KM607519	KM607226	KM607084	KM607371
KM607664	KM607793	KM607520	KM607227	KM607085	KM607372
KM607665	KM607794	KM607521	KM607228	KM607086	KM607373
KM607666	KM607795	KM607522	KM607229	KM607087	KM607374
KM607667	KM607796	KM607523	KM607230	KM607088	KM607375
AM418538		AM418566	AM418541	AM418569	AM418568
AY996562	GQ214699	EF593169	EU095948	EF520722	EF529519
AM418536		AM418540		AM418564	
FJ859722	FJ859748	FJ859735			
FJ859732	FJ859749	FJ859745			
FJ859723		FJ859736			
FJ859724		FJ859737			
FJ859725		Fj859738			
FJ859726		FJ859739			
FJ859727		FJ859740			
FJ859728		FJ859741			
FJ859729		FJ859742			
FJ859730		FJ859743			
FJ859731		FJ859744			
FJ859733	FJ859750	FJ859746			
FJ859734		FJ859747			
AM418534					
AM418535					
AM418537					
AM418539					
		AM418565			
		AM418567			
	AY996563				
			HE864318		
			HE864319		
					HE864320
JQ820459	JQ820460	JQ820461	JQ820462	JQ820463	JQ820464
JQ820465	JQ820466	JQ820467	JQ820468	JQ820469	JQ820470
			KC581796		
KM607600	KM607734	KM607453	KM607163	KM607019	KM607308
JF957628	JF957640	JF957652	JF957664	JF957676	JF957688
JF957625	JF957637	JF957649	JF957661	JF957673	JF957685
JF957626	JF957638	JF957650	JF957662	JF957674	JF957686
JF957636	JF957648	JF957660	JF957672	JF957684	JF957696
JF957635	JF957647	JF957659	JF957671	JF957683	JF957695
JF957627	JF957639	JF957651	JF957663	JF957675	JF957687
JF957632	JF957644	JF957656	JF957668	JF957680	JF957692
JF957634	JF957646	JF957658	JF957670	JF957682	JF957694
JF957633	JF957645	JF957657	JF957669	JF957681	JF957693
JF957629	JF957641	JF957653	JF957665	JF957677	JF957689
JF957630	JF957642	JF957654	JF957666	JF957678	JF957690
JF957631	JF957643	JF957655	JF957667	JF957679	JF957691
AF416467					
KM607655	KM607785	KM607511	KM607217	KM607075	KM607363
KM607669	KM607798	KM607525	KM607232	KM607090	KM607377
KM607670	KM607799	KM607526	KM607233	KM607091	KM607378
KM607671	KM607800	KM607527	KM607234	KM607092	KM607379
KM607683	KM607812	KM607539	KM607244	KM607103	KM607391
	KM607815	KM607542	KM607247	KM607106	KM607394
KM607686		KM607543	KM607248	KM607107	
KM607687	KM607816	KM607544	KM607249	KM607108	KM607395
KM607688	KM607817	KM607545	KM607250		KM607396

KF710622	KF710785	KF710948	KF11111	KF711274	KF711437			
KF710623	KF710786	KF710949	KF11112	KF711275	KF711438			
KF710624	KF710787	KF710950	KF11113	KF711276	KF711439			
KF710625	KF710788	KF710951	KF11114	KF711277	KF711440			
JX867550	KF435143	JX867540	JX569847	JX867548	JX867546	KF435145	KF435146	
JX867551								
	KF435144							
	KF435147							
		JX867541						
		JX867542						
		JX867543						
		JX867544						
		JX867545						
			KF435142					
				JX867549				
					JX867547			

KM607689 KM607818 KM607546 KM607251 KM607109 KM607397
KM607690 KM607819 KM607547 KM607252 KM607110 KM607398
KM607691 KM607820 KM607548 KM607253 KM607111 KM607399
KM607692 KM607821 KM607549 KM607254 KM607112 KM607400
KM607693 KM607822 KM607550 KM607255 KM607113 KM607401
KM607694 KM607823 KM607551 KM607256 KM607114 KM607402
KM607695 KM607552
KM607696 KM607824 KM607553 KM607257 KM607115 KM607403
KM607697 KM607825 KM607258 KM607116 KM607404
KM607698 KM607826 KM607259 KM607117 KM607405
KM607699 KM607827 KM607554 KM607260 KM607118 KM607406
KM607828 KM607555 KM607261 KM607119
KM607700 KM607556
KM607701 KM607557
KM607702 KM607829 KM607558 KM607262 KM607120 KM607407
KM607703 KM607830 KM607263 KM607121 KM607408
KM607704 KM607559 KM607264 KM607409
KM607410
KM607831 KM607560 KM607265 KM607122 KM607411
KM607705 KM607832 KM607561 KM607266 KM607123 KM607412
KM607706 KM607267 KM607124 KM607413
KM607833 KM607562 KM607268 KM607125 KM607414
KM607834 KM607563 KM607269 KM607126
KM607707 KM607835 KM607564 KM607270 KM607127 KM607415
KM607836 KM607565 KM607128 KM607416
KM607708 KM607837 KM607566 KM607271 KM607129
KM607709 KM607838 KM607567 KM607272 KM607130 KM607417
KM607839 KM607568 KM607273 KM607131 KM607418
KM607840 KM607569 KM607274 KM607132 KM607419
KM607710 KM607841 KM607570 KM607275 KM607133 KM607420
KM607842 KM607276 KM607134 KM607421
KM607843 KM607571 KM607277 KM607135 KM607422
KM607844 KM607278 KM607136
KM607711 KM607423
KM607572 KM607279 KM607137
KM607845 KM607573 KM607280 KM607138 KM607424
KM607712 KM607846 KM607281 KM607139 KM607425
KM607847 KM607426
KM607713 KM607282 KM607140
KM607714
KM607848 KM607574 KM607283 KM607141 KM607427
KM607849
KM607850 KM607284 KM607428
KM607851 KM607575 KM607285 KM607142 KM607429
KM607715 KM607852 KM607576 KM607286 KM607143 KM607430
KM607716 KM607853 KM607577 KM607287 KM607431
KM607717 KM607432
KM607854 KM607578 KM607288 KM607144 KM607433
KM607718 KM607855 KM607579 KM607289 KM607434
KM607856 KM607580 KM607290 KM607435
KM607719 KM607857 KM607581 KM607291 KM607145 KM607436
KM607720 KM607858 KM607582 KM607292 KM607146 KM607437
KM607583
KM607721 KM607859 KM607584 KM607293 KM607147 KM607438
KM607606 KM607742 KM607461 KM607027 KM607316
KM607463 KM607318
KM607608 KM607744 KM607465 KM607171 KM607028 KM607319
KM607610 KM607746 KM607468 KM607174 KM607031 KM607322
EF095161 EF095163 EF095164 EF095165 EF095166

DQ826390	DQ826391	DQ826393	DQ826394	DQ826395	DQ826396
EU366169	EU366170	EU366171	EU366172	EU366173	
KM607607	KM607743	KM607462			KM607317
		KM607464			
		AF148942			
AF416468					
	DQ826392				
EF095162					
	FJ773283				
	KM607788	KM607515			KM607368
KM607661	KM607789	KM607516	KM607223	KM607081	KM607369
KM607668	KM607797	KM607524	KM607231	KM607089	KM607376
KM607684	KM607813	KM607540	KM607245	KM607104	KM607392
KM607685	KM607814	KM607541	KM607246	KM607105	KM607393
KM607599	KM607733	KM607452	KM607162	KM607018	KM607307
U18077	U18078			U18079	
KM607660	KM607787	KM607514	KM607222	KM607080	KM607367
AB113659		AB113661			
AB113660		AB113662			
		AF148945			
AF416464					
AF416472					
AF416473					
AF416474					
AF416475					
AF416476					
AF416477					
AF416478					
AF416479					
KM607672	KM607801	KM607528	KM607235	KM607093	KM607380
					KM607381
KM607673	KM607802	KM607529	KM607236	KM607094	KM607382

Supplementary Table 4.2: Alphasatellite accession numbers of the sequences used in the CR-M and CR-SL analyses and the datasets they were divided into.

CR-M							
BBTV group 1	EU366174	AF216222	L32166	FJ394347	L32167	U02312	
BBTV group 2	GU074392	EU430730	U12587	U12586	FJ389724	AF416471	
	HQ616080	EU366175	AF216221	GU074391			
CBDV	KF435148						
CFDA	M29963						
Begomo group 1	AJ132344	FN554583	HF567947	FJ868830	EU384661	KF584010	EU384658
	AJ132345	FN658709	HG530544	FM164739	EU384662	KF584012	EU384659
	AJ512955	FN658727	HG530545	FM164740	EU384663	KJ028212	EU384660
	AJ512956	FN658728	HG530546	FM164939	EU589450	KJ843306	FN675300
	AJ512960	FN675284	HG530547	FN554580	FJ218494	KC305093	FN675301
	AJ512963	FN675285	HM004548	FN554581	FJ218495	KC305094	FN675302
	DQ641718	FN675286	HM446369	FN554582	FJ218496	KC763631	JX569789
	EU384652	FN675287	HQ180392	LK054802	FR717142	KC763632	HE984148
	EU384653	FN675288	HQ316180	GQ478667	FR772086	KC763633	HE984149
	EU384654	FN675289	HQ343234	GU992936	FR772087	KC763634	KM103524
	EU384655	FN675290	HQ728354	GU992937	FR873571	KC959931	GQ374450
	EU384656	FN675296	JF733780	HE858192	FR873573	JX183091	KC959932
	EU384657	FN675299	JX183090	HE858193	FR877532	JX262389	KF471047
Begomo group 2	AJ238493	DQ641717	FN658716	AJ579358	EU384627	HE599397	AM884370
	AJ512948	DQ641719	FN658717	AJ579359	EU384628	HF547408	AM930244
	AJ512949	EU384606	FN658718	AJ579361	EU384629	HF567944	AM930245
	AJ512950	EU384607	FN658729	AJ888445	EU384630	HG518788	AM930246
	AJ512951	EU384608	FN658730	AJ888446	EU384631	HG518789	AM930247
	AJ512952	EU384609	FN658735	AJ888447	EU384632	HG518790	AM930248
	AJ512954	EU384610	FN678899	AJ888448	EU384633	HG518791	EU384650
	AJ512957	EU384611	FN678900	AJ888449	EU384634	HG518792	EU384651
	AJ512958	EU384612	FN678901	AJ888450	EU384635	HQ407396	FJ218493
	AJ512959	EU384613	FN678902	AJ888451	EU384636	JQ322970	FM179614
	AJ579345	EU384614	FN678903	AJ888452	EU384637	JX570736	FN432360
	AJ579346	EU384615	FN806782	AJ888453	EU384638	JX913532	FN543100
	AJ579347	EU384616	FR772084	AJ888454	EU384639	KC282643	KM070823
	AJ579348	EU384617	FR772085	AJ888455	EU384640	KC305095	KM070824
	AJ579349	EU384618	FR772089	AM050734	EU384641	KC305096	KM070825
	AJ579350	EU384619	FR772090	AM050735	EU384642	KC677735	KM103523
	AJ579351	EU384620	FR772091	AM236763	EU384643	KC677736	KM103525
	AJ579352	EU384621	FR772092	AM236764	EU384644	KF267445	

CR-SL							
Babuvirus	KF435148	U12587	FJ389724	AF216221	U02312	L32167	
	AF416471	U12586	AF216222	GU074391	GU074392	FJ394347	
	EU430730	HQ616080	EU366174	EU366175	L32166		
CFDA	M29963						
Nanovirus alphasatellites	AJ005966	KC979008	KC978991	AB000921	KC978948		
	AJ132187	X80879	KC979051	AB009047	U16735		
	AJ005968	AJ132185	AB000922	KC979052	U16731		
	AJ005964	KC978990	AB000920	KC978957			
Geminivirus alphasatellites	AJ132344	EU384629	FR772090	EU384607	AJ888449	FJ218495	HQ180392
	AJ132345	EU384630	FR772091	EU384608	AJ888450	FJ218496	HQ316180
	AJ238493	EU384631	FR772092	EU384609	AJ888451	FJ868830	HQ343234
	AJ416153	EU384632	FR819709	EU384610	AJ888452	FJ956707	HQ407396
	AJ512948	EU384633	FR873571	EU384611	AJ888453	FM164739	HQ668024
	AJ512949	EU384634	FR873572	EU384612	AJ888454	FM164740	HQ728354
	AJ512950	EU384636	FR873573	EU384613	AJ888455	FM164939	JF733780
	AJ512951	EU384637	FR877532	EU384614	AM050734	FM179614	JQ041697
	AJ512952	EU384638	FR877533	EU384615	AM050735	FN432360	JQ322970
	AJ512954	EU384639	FR877536	EU384616	AM236763	FN436008	JX183090
	AJ512955	EU384640	GQ374450	EU384617	AM236764	FN543100	JX183091
	AJ512956	EU384641	GQ478667	EU384618	AM236765	FN554580	JX262389
	AJ512957	EU384642	GU385877	EU384619	AM236766	FN554581	JX569789
	AJ512958	EU384643	GU992936	EU384620	AM236767	FN554582	JX570736
	AJ512959	EU384644	GU992937	EU384621	AM749492	FN554583	JX913532
	AJ512960	EU384645	HE599396	EU384622	AM749493	FN658709	KC282643
	AJ512963	EU384646	HE599397	EU384623	AM749494	FN658711	KC305093
	AJ579345	EU384647	HE806451	EU384624	AM884370	FN658716	KC305094
	AJ579346	EU384648	HE858192	EU384625	AM930244	FN658717	KC305095
	AJ579347	EU384649	HE858193	EU384626	AM930245	FN658718	KC305096
	AJ579348	EU384650	HE984148	EU384627	AM930246	FN658727	KC577541
	AJ579349	EU384651	HE984149	EU384628	AM930247	FN658728	KC677735
	AJ579350	EU384652	HF547408	FN675286	AM930248	FN658729	KC677736
	AJ579351	EU384653	HF567944	FN675287	DQ641717	FN658730	KC763631
	AJ579352	EU384654	HF567947	FN675288	DQ641718	FN658735	KC763632
	AJ579353	EU384655	HG518788	FN675289	DQ641719	FN675284	KC763633
	AJ579354	EU384656	HG518789	FN675290	EU384606	FN675285	KC763634
	AJ579355	EU384657	HG518790	FN675296	FR717142	KF584011	KM103523

AJ579353	EU384622	FR819709	AM236765	EU384645	KF584011
AJ579354	EU384623	FR877533	AM236766	EU384646	KF584013
AJ579355	EU384624	FR877536	AM749492	EU384647	KF785752
AJ579356	EU384625	GU385877	AM749493	EU384648	KJ614231
AJ579357	EU384626	HE599396	AM749494	EU384649	KJ843304

AJ579356	EU384658	HG518791	FN675299	FR772084	KF584012	KM103524
AJ579357	EU384659	HG518792	FN675300	FR772085	KF584013	KM103525
AJ579358	EU384660	HG530543	FN675301	FR772086	KF785752	LK054802
AJ579359	EU384661	HG530544	FN675302	FR772087	KJ028212	NC023443
AJ579360	EU384662	HG530545	FN678899	FR772089	KJ614231	
AJ579361	EU384663	HG530546	FN678900	KC959931	KJ843304	
AJ888445	EU589450	HG530547	FN678901	KC959932	KJ843306	
AJ888446	FJ218492	HM004548	FN678902	KF267445	KM070823	
AJ888447	FJ218493	HM163578	FN678903	KF471047	KM070824	
AJ888448	FJ218494	HM446369	FN806782	KF584010	KM070825	

4.6 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Banerjee, A., Roy, S., Behere, G. T., Roy, S. S., Dutta, S. K. & Ngachan, S. V. (2014). Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. *Virus Res* 183, 41-49.

Basu, A. N. & Ganguly, B. (1968). A note on the transmission of foorkey disease of large cardamom by the aphid, *Micromyzus kalimpongensis* Basu. *Indian Phytopathology* 21, 127.

Beetham, P. R., Hafner, G. J., Harding, R. M. & Dale, J. L. (1997). Two mRNAs are transcribed from banana bunchy top virus DNA-1. *J Gen Virol* 78, 229-236.

Bell, K. E., Dale, J. L., Ha, C. V., Vu, M. T. & Revill, P. A. (2002). Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Arch Virol* 147, 695-707.

Blackman, R. L. & Eastop, V. P. (2008). *Aphids on the world's herbaceous plants and shrubs*. England: John Wiley and Sons Ltd.

Briddon, R. W., Bull, S. E., Amin, I., Idris, A. M., Mansoor, S., Bedford, I. D., Dhawan, P., Rishi, N., Siwatch, S. S., Abdel-Salam, A. M., Brown, J. K., Zafar, Y. & Markham, P. G. (2003). Diversity of DNA β, a satellite molecule associated with some monopartite begomoviruses. *Virology* 312, 106-121.

Briddon, R. W. & Stanley, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology* 344, 198-210.

Burns, T. M., Harding R. M. & Dale, J. L. (1995). The genome organisation of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76, 1471-1482.

Crooks, G., Hon, G., Chandonia, J. & Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Res* 14, 1188-1190.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-1797.

Fenoll, C., Schwarz, J. J., Black, D. M., Schneider, M. & Howell, S. H. (1990). The intergenic region of maize streak virus contains a GC-rich element that activates rightward transcription and binds maize nuclear factors. *Plant Mol Biol* 15, 865-877.

Fu, H. C., Hu, J. M., Hung, T. H., Su, H. J. & Yeh, H. H. (2009). Unusual events involved in banana bunchy top virus strain evolution. *Phytopathology* 99, 812-822.

Grigoras, I., Ginzo, A. I. d. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.-J. & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J Gen Virol* 95, 1178-1191.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59, 307-321.

Hafner, G. J., Harding, R. M. & Dale, J. L. (1997a). A DNA primer associated with banana bunchy top virus. *J Gen Virol* 78, 479-486.

Hafner, G. J., Stafford, M. R., Wolter, L. C., Harding, R. M. & Dale, J. L. (1997b). Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J Gen Virol* 78, 1795-1799.

Herrera-Valencia, V. A., Dugdale, B., Harding, R. M. & Dale, J. L. (2006). An iterated sequence in the genome of banana bunchy top virus is essential for efficient replication. *J Gen Virol* 87, 3409-3412.

Horser, C. L., Harding, R. M. & Dale, J. L. (2001a). Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *J Gen Virol* 82, 459-464.

Horser, C. L., Karan, M., Harding, R. M. & Dale, J. L. (2001b). Additional Rep-encoding DNAs associated with banana bunchy top virus. *Arch Virol* 146, 71-86.

Idris, A. M., Shahid, M. S., Briddon, R. W., Khan, A. J., Zhu, J. K. & Brown, J. K. (2011). An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. *J Gen Virol* 92, 706-717.

Islam, M. N., Naqvi, A. R., Jan, A. T. & Haq, Q. M. R. (2010). Genetic Diversity and Possible Evidence of Recombination among Banana Bunchy Top Virus (BBTV) Isolates. *International Research Journal of Microbiology* Vol 1, 001-012.

Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30, 772-780.

Katul, L., Timchenko, T., Gronenborn, B. & Vetten, H. J. (1998). Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *J Gen Virol* 79, 3101-3109.

King, A. M. Q., Lefkowitz, E., Adams, M. J. & Carstens, E. B. (2011). Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses. Burlington: Elsevier Science.

Magee, C. J. P. (1927). Investigation on the bunchy top disease of the banana. *Bulletin of the Council for Scientific and Industrial Research in Australia* 30.

Mandal, B., Mandal, S., Pun, K. B. & Varma, A. (2004). First report of the association of a nanovirus with foorkey disease of large cardamom in India. *Plant Dis* 88, 428-428.

Mandal, B., Shilpi, S., Barman, A. R., Mandal, S. & Varma, A. (2013). Nine novel DNA components associated with the foorkey disease of large cardamom: Evidence of a distinct babuvirus species in Nanoviridae. *Virus Res* 178, 297-305.

Mansoor, S., Khan, S. H., Bashir, A., Saeed, M., Zafar, Y., Malik, K. A., Briddon, R., Stanley, J. & Markham, P. G. (1999). Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* 259, 190-199.

Muhire, B. M., Varsani, A. & Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* 9, e108277.

Ndunguru, J., Legg, J. P., Fofana, I. B. F., Aveling, T. A. S., Thompson, G. & Fauquet, C. M. (2006). Identification of a defective molecule derived from DNA-A of the bipartite begomovirus of East African cassava mosaic virus. *Plant Pathology* 55, 2-10.

Patil, B. L. & Dasgupta, I. (2006). Defective interfering DNAs of plant viruses. *Crit Rev Plant Sci* 25, 47-64.

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25, 1253-1256.

Rey, M. E. C., Ndunguru, J., Berrie, L. C., Paximadis, M., Berry, S., Cossa, N., Nuaila, V. N., Mabasa, K. G., Abraham, N., Rybicki, E. P., Martin, D., Pietersen, G. & Esterhuizen, L. L. (2012). Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in Southern Africa, including the South-west Indian Ocean Islands. *Viruses* 4, 1753-1791.

Rohde, W., Randles, J. W., Langridge, P. & Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* 176, 648-651.

Saunders, K. & Stanley, J. (1999). A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: Evidence for interfamilial recombination between plant DNA viruses. *Virology* 264, 142-152.

Schubert, J., Habekuß, A., Wu, B., Thieme, T. & Wang, X. (2014). Analysis of complete genomes of isolates of the Wheat dwarf virus from new geographical locations and descriptions of their defective forms. *Virus Genes* 48, 133-139.

Sharman, M., Thomas, J. E., Skabo, S. & Holton, T. A. (2008). Abaca bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch Virol* 153, 135-147.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* 93, 1103-1119.

Stainton, D., Martin, D., Muhire, B., Lolohea, S., Halafihi, M., Lepoint, P., Blomme, G., Crew, K. S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D. A., Lemey, P., Harkins, G., Mabvakure, B., Thomas, J. E. & Varsani, A. (2015). The global distribution of Banana

bunch top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events *In Review*.

Stanley, J., Frischmuth, T. & Ellwood, S. (1990). Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. *Proc Natl Acad Sci U S A* 87, 6291-6295.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.

Timchenko, T., De Kouchkovsky, F., Katul, L., David, C., Vetten, H. J. & Gronenborn, B. (1999). A single Rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *J Virol* 73, 10173-10182.

Timchenko, T., Katul, L., Sano, Y., de Kouchkovsky, F., Vetten, H. J. & Gronenborn, B. (2000). The master Rep concept in nanovirus replication: Identification of missing genome components and potential for natural genetic reassortment. *Virology* 274, 189-195.

Varma, P. M. & Capoor, S. P. (1964). 'Foorkey' disease of large cardamom. *Indian J Agric Sci* 34, 56-62.

Vishnoi, R., Raj, S. & Prasad, V. (2009). Molecular characterization of an Indian isolate of Banana bunchy top virus based on six genomic DNA components. *Virus Genes* 38, 334-344.

Wang, H. I., Chang, C. H., Lin, P. H., Fu, H. C., Tang, C. & Yeh, H. H. (2013). Application of Motif-Based Tools on Evolutionary Analysis of Multipartite Single-Stranded DNA Viruses. *PLoS ONE* 8.

Watanabe, S., Greenwell, A. & Bressan, A. (2013). Localization, Concentration, and Transmission Efficiency of Banana bunchy top virus in Four Asexual Lineages of Pentalonia aphids. *Viruses* 5, 758-776.

Wu, R.-Y., You, L.-R. & Soong, T.-S. (1994). Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus. *Phytopathology* 84, 952-957.

Yu, N. T., Zhang, Y. L., Feng, T. C., Wang, J. H., Kulye, M., Yang, W. J., Lin, Z. S., Xiong, Z. & Liu, Z. X. (2012). Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* 44, 488-494.

Zhou, X. (2013). Advances in understanding begomovirus satellites. In *Annu Rev Phytopathol*, pp. 357-381.

Chapter 5

Identification of *Banana streak MY virus* (BSMYV), *Banana streak OL virus* (BSOLV) and *Banana streak GF virus* (BSGFV) banana-infecting badnaviruses in banana samples.

Contents

5.1. Abstract

Chlorotic and necrotic streaks on infected leaves are characteristics of Banana streak disease (BSD), a disease which leads to the eventual death of the infected plant. Rather than a single species being the causal agent of BSD, a number of banana-infecting badnavirus species have been identified. Banana-infecting badnaviruses (BIB) are able to exist in two forms, the episomal active form which causes disease in the host, and the endogenous form which is integrated into the host genome. These endogenous forms, however, can be reactivated and thus become episomal. Screening primers for three species of BIB, *Banana streak MY virus* (BSMYV), *Banana streak OL virus* (BSOLV) and *Banana streak GF virus* (BSGFV), were designed to amplify a region encompassing the movement protein domain of the polyprotein. These primers, along with previously designed primers were used to screen 267 worldwide banana samples, of which 82 were positive for at least one of the three species. The newly designed primers allowed better detection of BSMYV, BSOLV and BSGFV compared to primers previously used.

5.2. Introduction

BSD infects *Musa* spp. causing chlorotic leaf streaks which later become necrotic. These streaks run from the midrib to the leaf edge and the symptoms are not necessarily present across the entire leaf or even seen in all leaves of an infected plant (Daniells *et al.*, 2001; Lockhart, 1968). Along with the streak symptoms some infected plants also show leaf and bunch distortion and splitting of the pseudostem (Daniells *et al.*, 2001). BSD has been shown to delay banana harvest and consequently, a reduction in yield of up to 11% per annum, with the disease leading to eventual plant death (Daniells *et al.*, 2001; Lockhart, 1968). The disease is transmitted by a number of species of mealybugs including *Dysmicoccus brevipes*, *Planococcus ficus* and *Planococcus citri* (Geering *et al.*, 2005b; Lheureux *et al.*, 2007; Meyer *et al.*, 2008). Additionally, the disease is also spread by vegetative propagation, and through activation of the endogenous forms of the causal agent from the *Musa* genome.

BSD is caused by a number of BIB, all of which are members of the family *Caulimoviridae*, genus *Badnavirus*. There are currently four recognised species, *Banana streak MY virus* (BSMYV), *Banana streak OL virus* (BSOLV), *Banana streak GF virus* (BSGFV) and *Banana streak VN virus* (BSVNV) (Gayral *et al.*, 2008; Geering *et al.*, 2005b; Harper & Hull, 1998; Lheureux *et al.*, 2007). An additional six putative species, Banana streak UA virus (BSUAV), Banana streak UI virus (BSUIV), Banana streak UL virus (BSULV), Banana streak UM virus (BSUMV), Banana streak CA virus (BSCAV), and Banana streak IM virus (BSIMV) are yet to be officially classified as species (Geering *et al.*, 2011; James *et al.*, 2011b). In total, 27 full genomes of BIB have been sequenced and these are available in public databases.

Members of the badnaviruses contain an open circular double-stranded DNA genome (7.2-9.2kb) within a 30 x 130 nm bacilliform shaped virion (Lockhart, 1990; Lockhart, 1968). The genome contains three open reading frames (ORFs), with the polyprotein encoded on ORF 3 containing a movement protein motif, aspartic protease, reverse transcriptase and RNase H in this order (Chapter 1, Figure 1.7) (Geering *et al.*, 2005b; Harper & Hull, 1998). ORF 1 encodes a protein of unknown function whereas ORF 2 encodes a virion associated protein (Stavolone *et al.*, 2001).

BIB have also been found integrated into the Musa genome as endogenous sequences. The diversity seen in a number of endogenous sequences suggests multiple integration events have occurred (Geering *et al.*, 2005a). Although caulimoviruses replicate through reverse

transcription, they do not encode an integrase protein and replication is episomal, therefore integration into the host genome is thought to occur through the repair of DNA breaks or by recombination (Gayral *et al.*, 2008; Geering *et al.*, 2014; Hohn *et al.*, 2008). The majority of the integrated badnaviral sequences are thought to be defective, however, a number of endogenous BIB have been identified as activatable, even though the viral genomes are not intact (Chabannes *et al.*, 2013; Gayral *et al.*, 2008; Harper *et al.*, 1999a; Iskra-Caruana *et al.*, 2010; Ndowora *et al.*, 1999). Tissue culture of plants with endogenous forms of BIB has been shown to activate the virus (Côte *et al.*, 2010). As the majority of banana breeders rely on tissue culture, activatable endogenous BIB are a major issue for banana production.

Geering *et al.* (2000) designed diagnostic screening primers for BSOLV, BSGFV and BSMYV (RD-F1/R1, GF-F1/F2, Mys-F1/R1 respectively). All three primers are located at the 3′ end of ORF 3 and amplify a section of genome which includes the GRF zinc finger (Figure 5.1). These primers amplify both episomal and integrated forms of the virus species and have been used in a number of diagnostic studies, and have also been coupled with immunocapture PCR to detect only the episomal form of the virus (Geering *et al.*, 2000; James *et al.*, 2011a; Javer-Higginson *et al.*, 2014; Le Provost *et al.*, 2006; Sharma *et al.*, 2014).

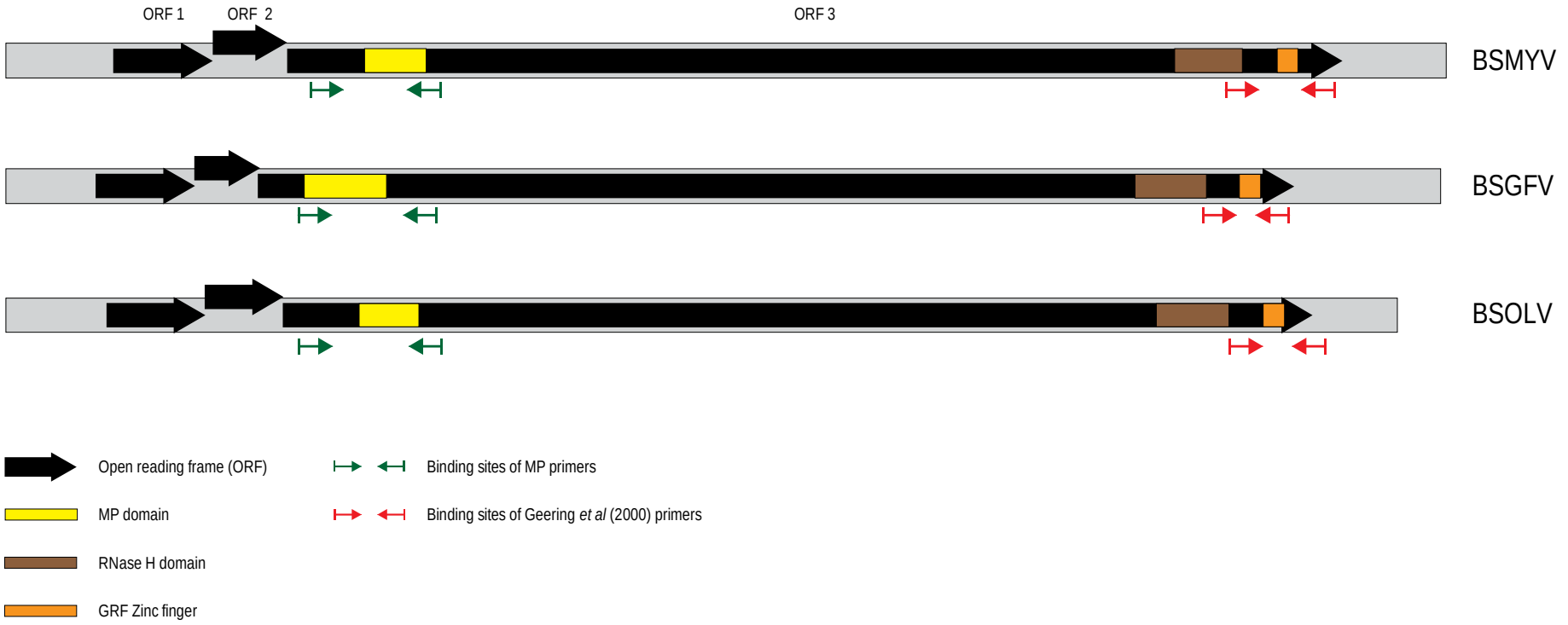


Figure 5.1: A cartoon illustration of BSMYV, BSGFV and BSOLV genomes with the location of the primer binding sites for all primers used in this study. Primer sequence information is available in Table 5.1. The three open reading frames are shown, along with the genome regions which are relevant to the primer binding sites, movement protein domain (MP domain), Ribonuclease H domain (RNase H domain) and the GRF Zinc finger.

5.3. Materials and Methods

Banana leaf material which was collected for banana bunchy top viral analysis (Chapters Two and Three) was also analysed for banana-infecting badnaviruses. The total DNA which was purified from these dried samples using the GenCatch Plant Genomic DNA Purification kit (Epoch Biolabs, USA), was used for PCR screening for banana-infecting badnaviruses.

5.3.1. Screening for endogenous badnavirus sequences

New screening primers Mys-MP-F/R, OL-MP-F/R and GF-MP-F/R were designed for BSMYV, BMOLV and BSGFV, respectively, from sequences available in GenBank. These primers bind at the 5′ of the ORF 3 polyprotein, encompassing the movement protein motif (MP), as identified in Pfam (Finn *et al.*, 2014) (Table 5.1; Figure 5.1). These novel primers, along with previously designed screening primers (Geering *et al.*, 2000) RD-F1/R1, GF-F1/F2 and Mys-F1/R1, which amplify the 3′ regions of ORF 3 BSOLV, BSGFV and BSMYV respectively (Table 5.1; Figure 5.1) (Geering *et al.*, 2000), were used to screen all the DNA extracted from the banana samples with KAPA 2G robust polymerase (KAPA Biosystems, USA). Thermocycling conditions for the screening primers OL-MP-F/R, RD-F1/R1, GF-F1/R1, MysF1/R1 and Mys-MP-F/R: 95°C for three minutes followed by 25 cycles of 95°C (15 sec), 53°C (15 sec), 72°C (15 sec) with a final extension of 72°C for two minutes. GF-MP-F/R required the same protocol but with an annealing temperature of 57°C. A selection of amplicons derived from the primer pairs which targeted the movement protein motif, designed as part of this study, were cloned and sequenced at Macrogen Inc (Korea).

5.3.2. Bioinformatic analyses

A maximum likelihood (ML) phylogenetic tree was constructed of the full genomes of the banana-infecting badnaviruses. Sequences were first aligned in MUSCLE (Edgar, 2004), with the best fit nucleotide model (GTR+G) determined in jModelTest (Posada, 2008) and constructed in PhyML 3 (Guindon *et al.*, 2010) with 100 bootstrap replicates, all implemented in MEGA 5 (Tamura *et al.*, 2011). The phylogenetic tree was rooted with Cauliflower mosaic virus (V00141) sequence and branches with <60% bootstrap support were collapsed using Mesquite v2.75 (http://mesquiteproject.org/).

Sequences from all three primers designed across the MP region OL-MP-F/R, Mys-MP-F/R and GF-MP-F/R were aligned with MUSCLE (Edgar, 2004) implemented in MEGA 5 (Tamura *et al.*, 2011) along with the equivalent regions of the full genomes as reference

sequences. A ML phylogenetic tree was constructed using PhyML 3 with HKY+G nucleotide substitution model selected using jModelTest, and 100 bootstrap replicates (branches with <60% support were collapsed). The ML phylogenetic MP tree was midpoint rooted.

Sequence Demarcation Tool (SDT) v1.2 (Muhire *et al.*, 2014), with the MUSCLE based alignment option, was used to determine the percentage pairwise nucleotide identities of the full genomes of the banana-infecting badnaviruses. Percentage pairwise identities were also determined for the sequenced screening primer MP amplicons designed in this study.

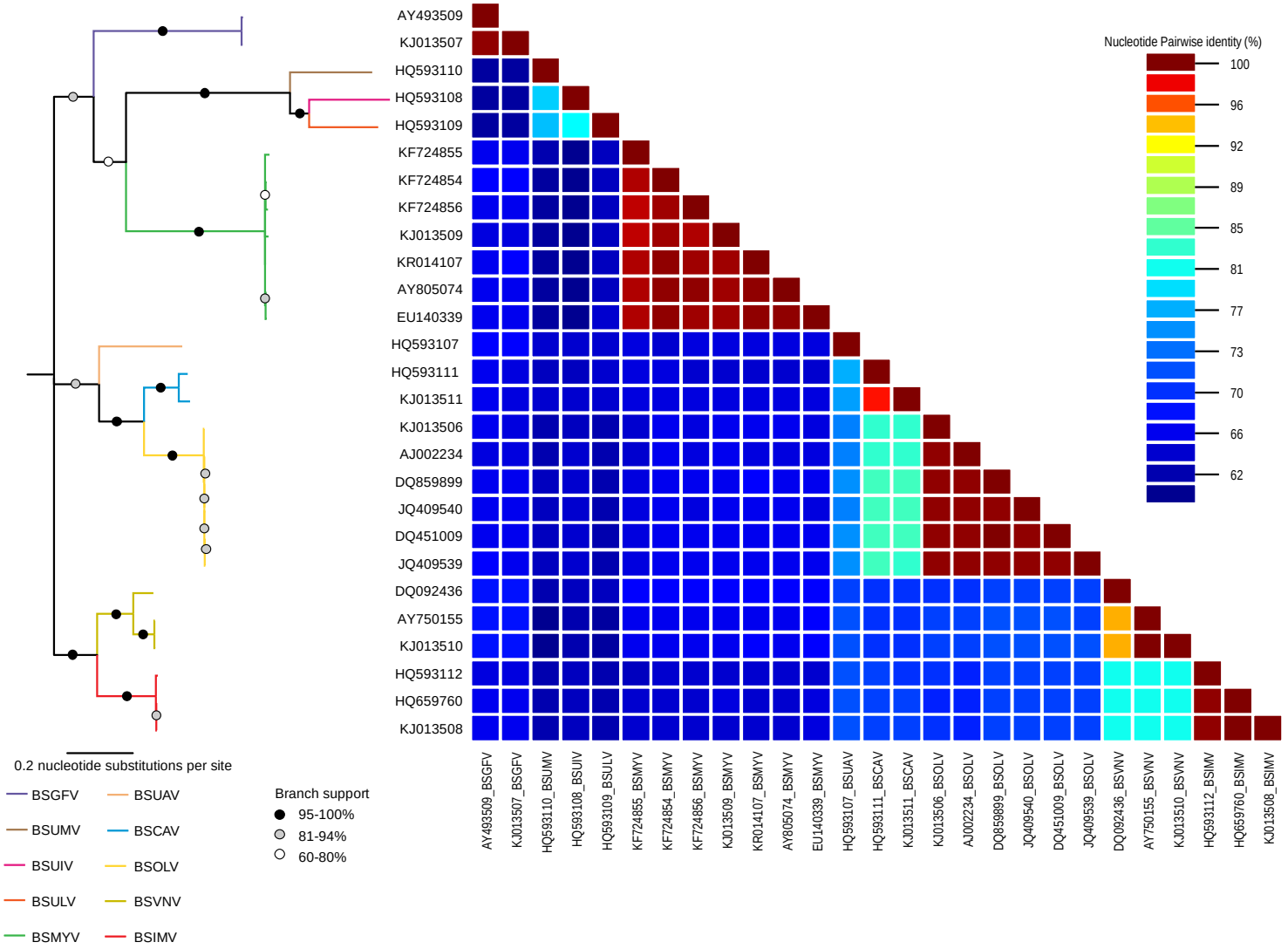


Figure 5.2 ML phylogenetic tree (nucleotide substitution model GTR+G) and percentage pairwise identity heatplot of the full genomes of banana-infecting badnaviruses. Branches with bootstrap support of <60% have been collapsed. Cauliflower mosaic virus sequence was used to root the phylogenetic tree

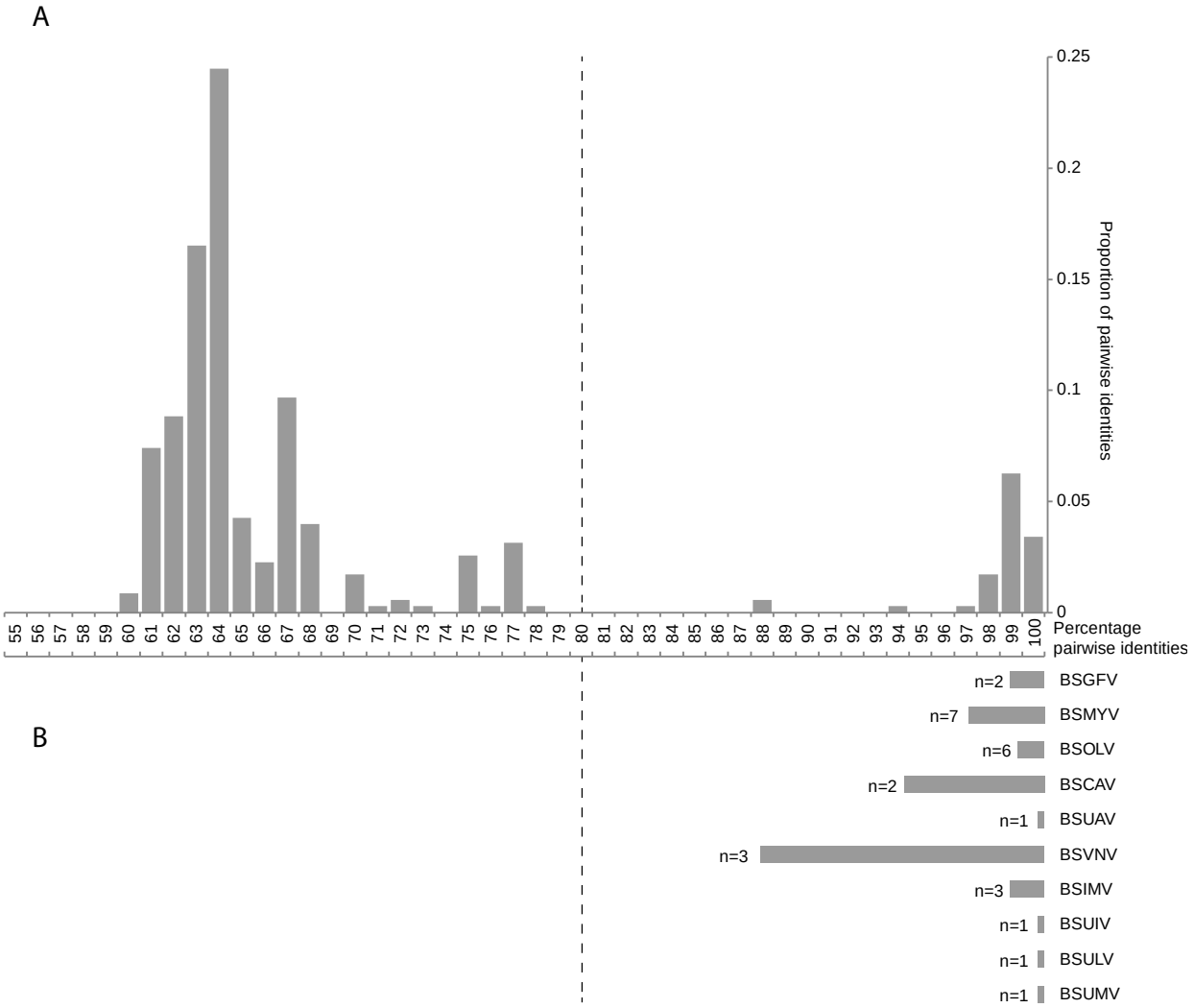


Figure 5.3: A) A distribution plot of the proportion of pairwise identity scores of all full genomes. The 80% species demarcation cutoff is shown with a dashed vertical line. B) Diversity of full genome isolates within each banana-infecting badnavirus species (or tentative species).

Table 5.1: Sequence information and amplicon length of all primers used in this study.

Primer name	Badnaviruses	Sequence	Amplicon size	Reference
Mys-F1	BSMYV	5'- TAAAAGCACAGCTCAGAACAAACC -3'	589 nt	Geering *et al* (2000)
Mys-R1	BSMYV	5'- CTCCGTGATTTCTTCGTGGTC -3'		
Mys-MP-F	BSMYV	5'- GGAGACGGTATGAAGCACAAC -3'	722 nt	This study
Mys-MP-R	BSMYV	5'- GCTTTGTGGCGTCGATAGC -3'		
RD-F1	BSOLV	5'- ATCTGAAGGTGTGTTGATCAATGC -3'	522 nt	Geering *et al* (2000)
RD-R1	BSOLV	5'- GCTCACTCCGCATCTTATCAGTC -3'		
OL-MP-F	BSOLV	5'- GAGCAGCATACGAAGCCCAAC -3'	708 nt	This study
OL-MP-R	BSOLV	5'- GATACTTTGTTGTACTCRGCTTCTG -3'		
GF-F1	BSGFV	5'- ACGAACTATCACGACTTGTTCAAGC -3'	476 nt	Geering *et al* (2000)
GF-R1	BSGFV	5'- TCGGTGGAATAGTCCTGAGTCTTC -3'		
GF-MP-F	BSGFV	5'- GTGCCCGATATGARGCCCAGAG -3'	705 nt	This study
GF-MP-R	BSGFV	5'- ACTTCATGGCGTCTATAGCCTTGATCC -3'		

5.4. Results and Discussion

5.4.1. Diversity of banana-infecting badnaviruses

The ML phylogenetic analysis of the full genomes of all BIB sequences, coupled with the pairwise identity analysis (Figure 5.2), clearly show the distinct species which are currently solely classified based on the RNase H gene, with <80% sequence similarity demarcating a new species (King *et al.*, 2011). These full genome analyses show (both phylogenetic and pairwise-identity based) that BIB are also able to be classified based on their full genomes, with <80% similarity demarcating a new species (Figure 5.3). BSVNV contains isolates with the greatest diversity with isolates sharing greater than 88% pairwise identity, with a full genome species demarcation not effecting these isolates (Figure 5.3B). A recent study has identified a number of recombination events within ORF 3 of the BIB and also the sugarcane-infecting badnaviruses (Sharma *et al.*, 2015), therefore species classification based on the full genome sequence is likely to be a more robust method than utilising a single region of the genome.

5.4.2. Screening of banana samples for banana-infecting badnaviruses

Two hundred and sixty seven banana leaf samples were screened for the presence of three species of banana-infecting badnaviruses, BSMYV, BSOLV and BSGFV using primer pairs designed in this study (Mys-MP-F/R; GF-MP-F/R; OL-MP-F/R) and those previously designed (Mys-F1/R1; GF-F1/R1; RD-F1/R1) by Geering *et al.* (2000). A stark difference was noticed in the detection of endogenous sequences of BSMYV, BSOLV and BSGFV using primers designed in this study and those by Geering *et al.* (2000) (Table 5.2). In the case of BSMYV, we obtained positive detection results for 36 samples using the Mys-F1/R1 primer pairs, however, with those designed in this study, we detected an additional 42 positives. In the case of BSOLV and BSGFV, 48 and 46 samples were positive with RD-F1/R1 and GF-F1/R1 primers and an additional 23 and six samples were positive with our primers respectively (Table 5.2).

Overall, 82 samples were found positive for BSMYV, BSOLV and / or BSGFV (Table 5.2). The positive samples, material sampled between 1989 and 2013, originate from 11 countries; Australia (n=6), Burundi (n=2), Democratic Republic of Congo (n=8), China (n=2), Egypt (n=1), India (n=5), Sri Lanka (n=2), Myanmar (n=1), the Philippines (n=3), Tonga (n=51) and Samoa (n=1) (Table 5.2). BSMYV was the most prevalent banana-infecting badnavirus in the samples tested with 78 positive for BSMYV, 72 for BSOLV and 53 for BSGFV. Forty

seven samples were found to be positive for BSMYV, BSOLV and BSGFV, 26 samples positive for two species, and only nine samples positive for a single species (Table 5.2).

The BSMYV, BSOLV and BSGFV screening primers could be detecting episomal or integrated viral sequences. Nonetheless, given that the integrated sequences can become activated, (Gayral *et al.*, 2008; Harper *et al.*, 1999b; Iskra-Caruana *et al.*, 2010; Ndowora *et al.*, 1999) detection of both integrated or episomal, it is essential to identify these for banana breeding programs.

Amplicons derived from Mys-MP-F/R (n=72), GF-MP-F/R (n=39) and OL-MP-F/R (n=42) primer pairs were sequenced to determine the specificity of the primers and identify potential diversity in these sequences. The equivalent regions of the full genome sequences of banana infecting badnaviruses, including those of BSMYV, BSOLV and BSGFV were also identified and included in the subsequent data analysis. The ML phylogenetic tree shows all amplicons resulting from the BSMYV, BSOLV and BSGFV specific primers are species-specific (Figure 5.4). The BSMYV sequences showed much higher diversity than either BSOLV or BSGFV with no geographical clustering pattern seen in the different BSMYV groups. Overall the BSMYV MP region sequences share >83.5% pairwise identity, whereas the BSOLV sequences share >93.7% and the BSGFV share >98% pairwise identity.

Table 5.2: Banana material which was positive (+) for the banana-infecting badnaviruses; BSMYV (Mys-F1/R1, Mys-Mp-F/R), BSGFV (GF-F1/R1, GF-Mp-F/R) and / or BSOLV (RD-F1/R1, OL-Mp-F/R). Where BBTV isolates have previously been determined for these samples (Chapter 2, Chapter 3), the phylogenetic group of BBTV isolates are given, SPG – South Pacific phylogenetic group, AG – Asian phylogenetic group. Banana sample information, including year of collection and country of collection: AU-Australia, BI-Burundi, CD-Democratic Republic of Congo, CN-China, EG-Egypt, IN-India, LK-Sri Lanka, MM-Myanmar, PH-Philippines, TO-Tonga, WS-Samoa.

Country	Study code	Year	BBTV sequenced	Mys-F1/R1	Mys-Mp-F/R	GF-F1/R1	GF-Mp-F/R	RD-F1/R1	OL-Mp-F/R
AU	B2845	2011	SPG		+				+
AU	KP17	2010	SPG	+	+				+
AU	KP18	2010	SPG	+	+				+
AU	602	1996	SPG		+				
AU	737	1997	SPG		+				
AU	2557	2010	SPG	+	+				+
BI	526	1992	SPG		+	+	+	+	+
BI	547	1995	SPG	+	+				+
CD	BU06	2012	SPG	+	+	+	+	+	+
CD	BU08	2012		+	+			+	
CD	BU09	2012	SPG	+	+	+	+	+	+
CD	BU10	2012	SPG	+	+	+	+	+	+
CD	BU11	2012	SPG	+	+	+	+		
CD	BU12	2012	SPG	+	+	+	+	+	+
CD	BU15	2012	SPG	+	+	+	+	+	+
CD	BU16	2012	SPG		+	+	+		+
CN	Q529-4	1990	AG		+				
CN	Q529-3	1990			+				
EG	9-150510	2010	SPG		+	+	+	+	+
IN	523-6A	1991	SPG/AG		+				
IN	523-6B	1991	SPG/AG		+				
IN	Q524-1		SPG	+	+				+
IN	Q524-2		SPG	+	+				
IN	Q524-3		SPG		+				
LK	KP5	2003	SPG	+	+				+
LK	Q553	1995	SPG		+				+
MM	Daff-24720	2013			+	+	+	+	+
PH	MS6	2008	AG	+	+	+	+	+	+
PH	571-1	1993	AG	+	+			+	+
PH	571-2	1993	AG	+	+	+	+	+	+
TO	TOS6	2010		+	+	+	+	+	+
TO	TOS7	2010	SPG	+	+	+	+	+	+
TO	TOS8	2010		+	+	+	+	+	+
TO	TOS10	2010		+	+				+
TO	TOS11	2010			+	+	+	+	+
TO	TOS12	2010	SPG		+	+	+	+	+
TO	TOS16	2010	SPG		+	+	+	+	+

TO	TOS20	2010	SPG			+	+	+	+	+
TO	TOS21	2010	SPG			+	+	+	+	+
TO	TOS22	2010	SPG			+		+	+	+
TO	TOS23	2010			+	+		+	+	+
TO	TOS24	2010			+	+	+	+	+	+
TO	TOS25	2010	SPG			+	+	+	+	+
TO	TOS26	2010					+	+	+	+
TO	TOS28	2010	SPG			+	+	+	+	+
TO	TOS29	2010	SPG				+	+		+
TO	TOS31	2010				+		+		+
TO	TOS32	2010				+				+
TO	TOS33	2010				+				+
TO	TOS39	2010	SPG			+		+		+
TO	TOS40	2010	SPG		+	+				+
TO	TOS41	2010			+	+	+	+	+	+
TO	TOS44	2010			+	+	+	+	+	+
TO	TOS47	2010				+	+	+	+	+
TO	TOS48	2010	SPG			+		+	+	+
TO	TOS56	2010	SPG			+	+	+	+	+
TO	TOS62	2010	SPG			+	+	+	+	+
TO	TOS88	2010	SPG		+	+	+	+	+	+
TO	TO217	2010			+	+				+
TO	TO218	2010			+	+				+
TO	TO220	2010			+	+				+
TO	TO113	2010				+	+	+	+	+
TO	TO114	2010	SPG			+	+	+	+	+
TO	TO121	2010	SPG			+	+	+	+	+
TO	TO164	2010				+	+	+	+	+
TO	TO166	2010	SPG			+	+	+	+	+
TO	TO208	2010			+	+				+
TO	TO210	2010			+	+				
TO	TO213	2010			+	+				+
TO	TO214	2010			+	+				+
TO	TO224	2010	SPG				+	+	+	+
TO	TO226	2010				+	+	+	+	+
TO	TO265	2010				+	+	+	+	+
TO	TO266	2010				+	+	+	+	+
TO	TO268	2010				+	+	+	+	+
TO	TO270	2010				+	+	+	+	+
TO	TO303	2010				+	+	+	+	+
TO	TO314	2010	SPG			+	+	+	+	+
TO	KP4	1990	SPG		+	+	+	+	+	+
TO	Q276	1989	SPG				+	+	+	+
TO	Q570	1990	SPG			+		+		+
WS	Q281	1989	SPG		+	+				+

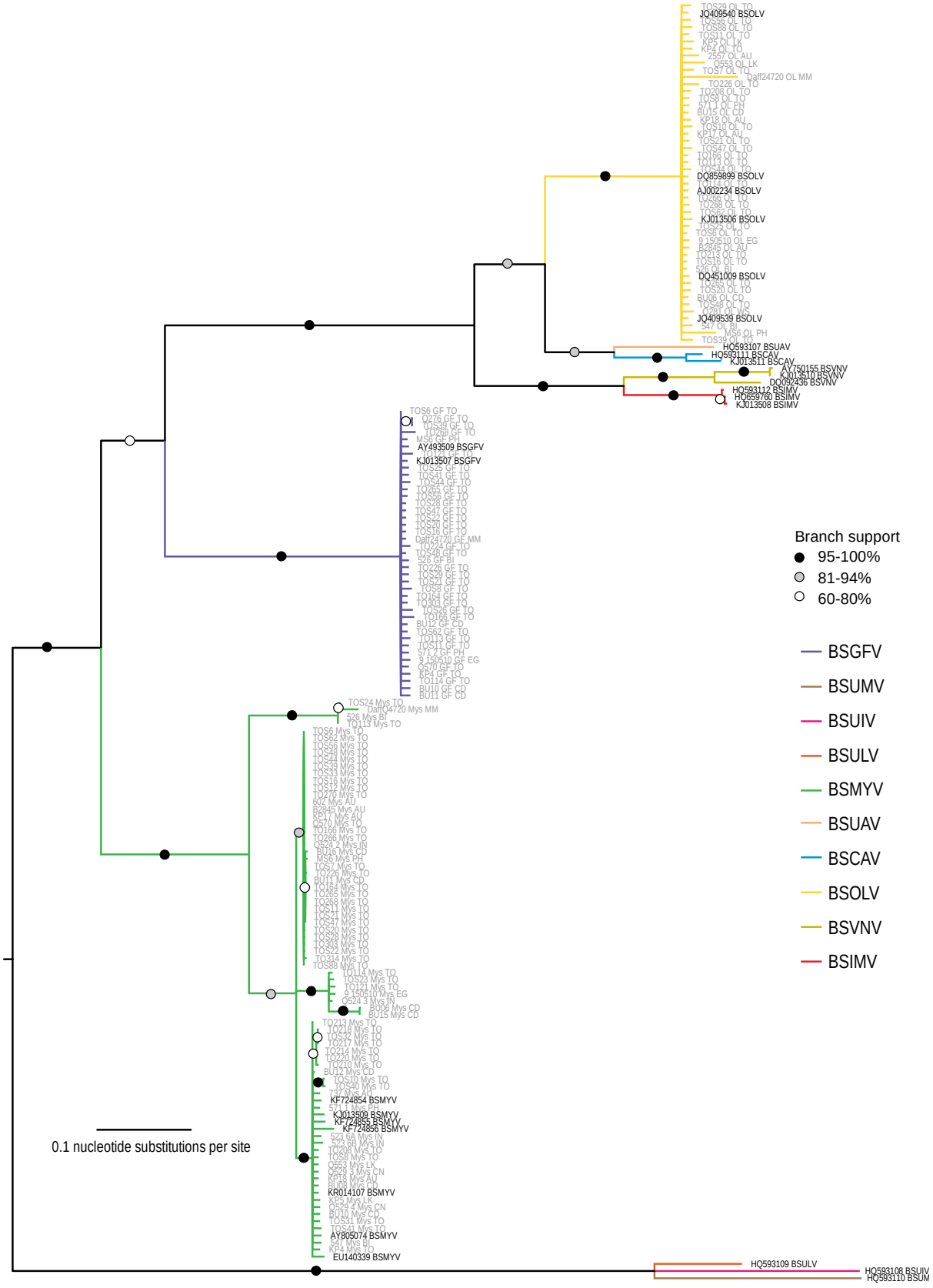


Figure 5.4: ML phylogenetic tree of the MP amplicons of banana-infecting badnaviruses inferred using nucleotide substitution model HKY+G. Equivalent regions from the full genome sequences are included with accession numbers shown in bold. The phylogenetic tree is midpoint rooted and branches with <60% bootstrap support have been collapsed.

5.4.3. Badnaviruses and *Banana bunchy top virus* (BBTV)

Of the 82 samples which were positive for banana-infecting badnaviruses, in 51 we also had identified at least one component of BBTV (Chapter Two and Chapter Three) (Table 5.2). BBTV component sequences fall into two phylogenetic groups defined as the South Pacific group (SPG) and Asian group (AG). Forty-five samples contained BBTV components of the SPG (Australia n=6, Burundi n=2, Democratic Republic of Congo n=7, Egypt n=1, India n=3, Sri Lanka n= 2, Tonga n=23 Samoa n=1), two samples contained components which are found in both the South Pacific and Asian phylogenetic groups (SPG/AG) (India n=2), and four samples contained BBTV components which were found in the AG (China n=1, Philippines n=3).

The vast majority of samples which were infected with BBTV and a badnavirus, contained BSMYV (94%), with only three samples from Tonga not containing BSMYV, instead containing BSGFV and / or BSOLV. No apparent trends between badnavirus species and the BBTV phylogenetic groups were identified, however, country specific trends were evident. Of note a large number of Tongan samples (n=19) contained BBTV, BSMYV, BSGFV and BSOLV, all the Democratic Republic of Congo samples with BBTV also contained both BSMYV and BSGFV with 6/7 samples also containing BSOLV and all the samples from the Philippines contained BSMYV, BSOLV and 2/3 also contained BSGFV indicating that all three badnaviruses are prevalent in these countries. Interestingly no samples from Australia, Sri Lanka or India contained BSGFV, suggesting these countries may be free of this badnavirus species.

5.5. Concluding remarks

Screening primers were designed for three banana-infecting badnaviruses, BSMYV, BSOLV and BSGFV which encompass the movement protein motif located on ORF 3. These primers, are species-specific and were found to be more efficient at detecting BSMYV, BSOLV and BSGFV sequences in banana material than previously designed sets of primers by Geering *et al.* (2000). BIB were detected in 82 samples collected in 11 countries, with the majority of positive samples containing more than one badnavirus species, with BSMYV the most prevalent.

Other than a single Myanmar sample and a number of Tongan samples (n=28), all other samples (n=51) with identified BSMYV, BSOLV and / or BSGFV have also had BBTV

components sequenced indicating mixed infections. Four samples contained BBTV components in the AG group, two samples contained BBTV components in both the SPG and AG groups and 45 samples contained SPG BBTV.

Due to the activateable nature of these viruses, screening for both the integrated or episomal forms of BSMYV, BSOLV and BSGFV is important. The screening primers designed in this study across the MP region of ORF 3 show higher detection rates than primers designed in a different region of the ORF 3, making these novel primers an important diagnostic tool.

5.6. References

Chabannes, M., Baurens, F. C., Duroy, P. O., Bocs, S., Vernerey, M. S., Rodier-Goud, M., Barbe, V., Gayral, P. & Iskra-Caruana, M. L. (2013). Three infectious viral species lying in wait in the banana genome. *J Virol* 87, 8624-8637.

Côte, F. X., Galzi, S., Folliot, M., Lamagnère, Y., Teycheney, P. Y. & Iskra-Caruana, M. L. (2010). Micropropagation by tissue culture triggers differential expression of infectious endogenous Banana streak virus sequences (eBSV) present in the B genome of natural and synthetic interspecific banana plantains. *Mol Plant Pathol* 11, 137-144.

Daniells, J. W., Geering, A. D. W., Bryde, N. J. & Thomas, J. E. (2001). The effect of Banana streak virus on the growth and yield of dessert bananas in tropical Australia. *Ann Appl Biol* 139, 51-60.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-1797.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. & Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Res* 42, D222-D230.

Gayral, P., Noa-Carrazana, J. C., Lescot, M., Lheureux, F., Lockhart, B. E. L., Matsumoto, T., Piffanelli, P. & Iskra-Caruana, M. L. (2008). A single Banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. *J Virol* 82, 6697-6710.

Geering, A. D. W., Maumus, F., Copetti, D., Choisne, N., Zwickl, D. J., Zytnicki, M., McTaggart, A. R., Scalabrin, S., Vezzulli, S., Wing, R. A., Quesneville, H. & Teycheney, P.-Y. (2014). Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nat Commun* 5.

Geering, A. D. W., McMichael, L. A., Dietzgen, R. G. & Thomas, J. E. (2000). Genetic diversity among banana streak virus isolates from Australia. *Phytopathology* 90, 921-927.

Geering, A. D. W., Olszewski, N. E., Harper, G., Lockhart, B. E. L., Hull, R. & Thomas, J. E. (2005a). Banana contains a diverse array of endogenous badnaviruses. *J Gen Virol* 86, 511-520.

Geering, A. D. W., Parry, J. N. & Thomas, J. E. (2011). Complete genome sequence of a novel badnavirus, banana streak IM virus. *Arch Virol* 156, 733-737.

Geering, A. D. W., Pooggin, M. M., Olszewski, N. E., Lockhart, B. E. L. & Thomas, J. E. (2005b). Characterisation of banana streak mysore virus and evidence that its DNA is integrated in the B genome of cultivated Musa. *Arch Virol* 150, 787-796.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59, 307-321.

Harper, G., Dahal, G., Thottappilly, G. & Hull, R. (1999a). Detection of episomal banana streak badnavirus by IC-PCR. *J Virol Methods* 79, 1-8.

Harper, G. & Hull, R. (1998). Cloning and sequence analysis of banana streak virus DNA. *Virus Genes* 17, 271-278.

Harper, G., Osuji, J. O., Heslop-Harrison, J. S. & Hull, R. (1999b). Integration of banana streak badnavirus into the Musa genome: Molecular and cytogenetic evidence. *Virology* 255, 207-213.

Hohn, T., Richert-Pöggeler, K., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C., Teycheney, P.-Y., Iskra-Caruana, M.-L. & Hull, R. (2008). Evolution of Integrated Plant Viruses. In *Plant Virus Evolution*, pp. 53-81. Edited by M. Roossinck: Springer Berlin Heidelberg.

Iskra-Caruana, M. L., Baurens, F. C., Gayral, P. & Chabannes, M. (2010). A four-partner plant-virus interaction: Enemies can also come from within. *Mol Plant-Microbe Interact* 23, 1394-1402.

James, A. P., Geijskes, R. J., Dale, J. L. & Harding, R. M. (2011a). Development of a novel rolling-circle amplification technique to detect Banana streak virus that also discriminates between integrated and episomal virus sequences. *Plant Dis* 95, 57-62.

James, A. P., Geijskes, R. J., Dale, J. L. & Harding, R. M. (2011b). Molecular characterisation of six badnavirus species associated with leaf streak disease of banana in East Africa. *Ann Appl Biol* 158, 346-353.

Javer-Higginson, E., Acina-Mambole, I., González, J. E., Font, C., González, G., Echemendía, A. L., Muller, E. & Teycheney, P. Y. (2014). Occurrence, prevalence and molecular diversity of banana streak viruses in Cuba. *Eur J Plant Pathol* 138, 157-166.

King, A. M. Q., Lefkowitz, E., Adams, M. J. & Carstens, E. B. (2011). Virus Taxonomy : Ninth Report of the International Committee on Taxonomy of Viruses. Burlington: Elsevier Science.

Le Provost, G., Iskra-Caruana, M.-L., Acina, I. & Teycheney, P.-Y. (2006). Improved detection of episomal Banana streak viruses by multiplex immunocapture PCR. *J Virol Methods* 137, 7-13.

Lheureux, F., Laboureau, N., Muller, E., Lockhart, B. E. L. & Iskra-Caruana, M. L. (2007). Molecular characterization of banana streak acuminata Vietnam virus isolated from Musa acuminata siamea (banana cultivar). *Arch Virol* 152, 1409-1416.

Lockhart, B. (1990). Evidence for a Double-Stranded Circular DNA Genome in a Second Group of Plant Viruses. *Phytopathology* 80, 127-131.

Lockhart, B. E. L. (1968). Purification and serology of a bacilliform virus associated with banana streak disease. *Phytopathology* 76, 995-999.

Meyer, J. B., Kasdorf, G. G. F., Nel, L. H. & Pietersen, G. (2008). Transmission of activated-episomal Banana streak OL (badna)virus (BSOLV) to cv. Williams banana (Musa sp.) by three mealybug species. *Plant Dis* 92, 1158-1163.

Muhire, B. M., Varsani, A. & Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* 9, e108277.

Ndowora, T., Dahal, G., Lafleur, D., Harper, G., Hull, R., Olszewski, N. E. & Lockhart, B. (1999). Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. *Virology* 255, 214-220.

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25, 1253-1256.

Sharma, S. K., Vignesh Kumar, P., Geetanjali, A. S., Pun, K. B. & Baranwal, V. K. (2015). Subpopulation level variation of banana streak viruses in India and common evolution of banana and sugarcane badnaviruses. *Virus Genes*.

Sharma, S. K., Vignesh Kumar, P., Poswal, R., Rai, R., Swapna Geetanjali, A., Prabha, K., Jain, R. K. & Baranwal, V. K. (2014). Occurrence and distribution of banana streak disease and standardization of a reliable detection procedure for routine indexing of banana streak viruses in India. *Scientia Horticulturae* 179, 277-283.

Stavolone, L., Herzog, E., Leclerc, D. & Hohn, T. (2001). Tetramerization is a conserved feature of the virion-associated protein in plant pararetroviruses. *J Virol* 75, 7739-7743.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.

Chapter 6

Conclusion

Contents

6.1 Overview

Bananas are an important commercial and subsistence crop grown in over 130 countries (FAOSTAT, 2012) around the world. The word banana encompasses all banana fruit, both the dessert and cooking varieties, with banana the main starch crop for millions of people. The majority of banana cultivars are the result of the crossing of diploid wild bananas, *Musa acuminata* and *Musa balbisiana*, resulting in sterile triploid cultivars. *M. balbisiana* and *M. acumina* are naturally found in the Malaysian/Indonesian/Philippines region, with evidence of early agriculture in Papua New Guinea (6950-6440 BP) and Cameroon (2750-2300 BP) (Denham *et al.*, 2004; Mbida Mindzie *et al.*, 2001; Perrier *et al.*, 2011). Sterile banana cultivars are propagated vegetatively through the removal of suckers which form at the base of the trunk, and more recently through tissue culture. Both *Banana bunchy top virus* (BBTV) and banana-infecting badnaviruses (BIB) infect *Musa* spp. resulting in yield loss and often total plant death, with both named as economically important viruses of banana (Rybicki, 2014). Outbreaks of BBTV in particular cause large scale crop loss.

Since BBTV has a multi-component genome, like other nanoviruses, these viruses are able to evolve through reassortment and recombination (Banerjee *et al.*, 2014; Grigoras *et al.*, 2014; Hu *et al.*, 2007; Hyder *et al.*, 2011; Savory & Ramakrishnan, 2014; Stainton *et al.*, 2012; Stainton *et al.*, 2015; Wang *et al.*, 2013; Yu *et al.*, 2012). To improve the resolution of detection of recombination and reassortment events, an increase in the number of component sequences is required, and in the case of reassortment an increase in the number of full genomes is preferential. At the beginning of this thesis only five full BBTV genomes were available (Australia n=1, India n=1, Taiwan n=1, China n=2).

BBTV components fall into two phylogenetically defined groups, the South Pacific Group (SPG) and the Asian Group (AG), which were first determined using DNA-R components from ten countries. Generally BBTV component sequences fall into these two phylogenetic groups, even with the addition of isolates from different countries. All BBTV components contain two common regions, these are involved in recognition and replication by the replication-associated protein, and the initiation of secondary strand synthesis.

Multiple species of BIB are the causal agents of banana streak disease. These pararetroviruses are able to persist both in an episomal infectious state and in an endogenous state within the *Musa* genome. The endogenous form has the potential to activate into the episomal form and

cause infection (overview in Chapter One). Therefore, detection of these viruses is paramount to ensuring clean banana stock.

This PhD thesis aimed to further investigate the global genetic diversity of BBTV including the characterisation of the common regions, investigate the evolution of BBTV through identifying evidence of recombination and reassortment events and to use geographical structuring to determine potential large-scale movement of BBTV. The presence of three species of BIB are also investigated in globally sourced banana samples.

6.2 Major findings

6.2.1 Summary

In total, 927 components of BBTV were sequenced from 183 banana samples (a total of 267 samples were screened for BBTV) as part of this thesis research. These components include 106 full genomes and these sequences were analysed along with all full component sequences available in GenBank in order to gain insights into the evolutionary dynamics of BBTV (Chapter Two and Three). The common regions of the sequences from this large BBTV dataset, together with publically available sequences of the two other babuvirus species, *Abaca bunchy top virus* (ABTV) and *Cardamom bushy dwarf virus* (CBDV), and alphasatellite sequences, were analysed (Chapter Four). Twenty four defective molecules, which were identified during BBTV component sequencing (Chapter Three), were characterised in Chapter Four. The defective molecules all contain insertions and / or deletions which in the majority of sequences resulted in a disrupted open reading frame (ORF).

Banana samples were also screened for the presence of three banana-infecting badnavirus species, *Banana streak MY virus* (BSMYV), *Banana streak OL virus* (BSOLV) and *Banana streak GF virus* (BSGFV). Novel species specific primers were designed which detected a greater number of badnaviruses than previously designed primers (Geering *et al.*, 2005). In total 267 banana samples were screened with both the novel and previously designed primers with 82 samples positive for at least one badnavirus species. Of these 82 samples, 51 have also had BBTV components sequenced as part of this study.

6.2.2 Diversity and evolution of BBTV

A preliminary study was undertaken to gather insights into the evolutionary mechanisms of BBTV (Chapter Two). Samples were collected from Tonga and full genomes of 12 isolates (72 components) were recovered, sequenced and analysed with all available BBTV components in GenBank. These were the first BBTV genomes sequenced from Tongan samples. The research within Chapter Two identified that both recombination and reassortment are common in BBTV evolution. Concatenated genomes allowed the use of the Recombination Detection Programme (RDP4.27) to detect reassorted genomes, where breakpoints which spanned entire components were indicative of reassortment. This method has subsequently been utilised for other nanovirus genomes to detect reassortment events (Grigoras *et al.*, 2014; Savory & Ramakrishnan, 2014).

Subsequent to the initial study, a larger worldwide study was undertaken where a further 855 components were sequenced (DNA-R, n=137; DNA-U3, n=138; DNA-S, n=146; DNA-M, n=146; DNA-C, n=143; DNA-N, n=145) from 171 banana samples collected from 14 countries (Chapter Three). In this study, 94 full genomes were determined including those from Burundi, Democratic Republic of Congo, Egypt, Indonesia, the Philippines and Hawaii which had disparate genomes available but no full genomes. Full genomes from Congo and Samoa were also sequenced for which there was no BBTV sequence data previously. The increase in sequence data allowed for more robust analyses into the evolutionary dynamics of BBTV. All component sequences from Chapter Two and Three, and those available on GenBank, were compiled into datasets (total components analysed 1,191; DNA-R: n=242; DNA-U3: n=190; DNA-S: n=225; DNA-M: n=186; DNA-C: n=180; DNA-N: n=181) for phylogenetic and evolutionary analyses. A selection analysis of all BBTV and CBDV genes identified that both species were evolving under predominantly negative selection, with individual sites across all genes also evolving under positive selection. Unsurprisingly, due to the conserved nature of the Rep and Clink motifs, all are evolving under negative or neutral selection, with one exception, a single site in the Walker B motif within the Rep of BBTV. Percentage pairwise identities of all isolates identified greater genetic diversity within the AG genome components than the SPG for all components except DNA-S. All components were further split into five major regions of the world, where the Southeast Asia/Far East and the Indian subcontinent regions were found to harbour the greatest diversity overall.

As expected, the large increase in component sequences, and in particular the increase in full genomes, resulted in an increase of detected reassortment events, with ~38% of isolates with at least three components available containing evidence of reassortment events, similar to what was seen in CBDV (Savory & Ramakrishnan, 2014). Since multiple reassortment events were detected in multiple genomes, it is likely these events occurred in ancestral genomes and have been retained and continue to circulate in the population. Savory and Ramakrishnan (2014) identified DNA-M and DNA-N in CBDV as the most reassorted components, however, in BBTV DNA-U3 was found to be the most reassorted component suggesting species or host differences in reassortment rather than function. The genome formula of number of relative components present in the host, which was recently determined for a nanovirus (Sicard *et al.*, 2013), may also play a role in which components are more likely to be reassorted due to the relative number of each component available, however, the genome formula has not be determined for either BBTV or CBDV.

The component DNA-U3 also contained the greatest number of detected recombination events. As was seen in the reassortment analysis, the increase in the number of full components also resulted in an increase in detectible recombination events, 34 events detected with 12 of those within DNA-U3. In Chapter Two we identified two inter-recombination hotspots, one across each of the two common regions, the common region stem-loop (CR-SL) and common region major (CR-M). In Chapter Three due to the increase in sequences and issues of aligning the different components, it was more credible to identify inter-recombination from analysis of unknown recombinant regions identified in the intra-component datasets. Therefore inter-recombination hotspots were unable to be determined in Chapter Three. However, in Chapter Three, all components had evidence of recombination involving the CR-SL, with recombination events involving the CR-M only identified in DNA-N and DNA-U3, suggesting that the CR-SL is a hotspot for recombination for all components.

As the total number of available BBTV genomes had increased to 121 genomes, we were able to analyse the geographical structure of BBTV (Chapter Three). Recombinant regions and reassorted components were removed from the concatenated full genome sequences in order to construct a recombination-free phylogenetic tree. Geographical clustering of isolates was evident across this full genome phylogenetic tree, with monophyletic clades indicating single founder events for Hawaii, Tonga and Australia. The Hawaiian isolates are located

within the clade of Samoan isolates suggesting that these isolates originated from Samoa. Although there is evidence of movement within regions, with intermingling seen between samples from Taiwan, Indonesia and China, there is little evidence of frequent large continental movements.

6.2.3 Common regions of BBTV

An increase in the number of BBTV components (Chapter Two and Three), as well the recent deposit of a large CBDV dataset (Savory & Ramakrishnan, 2014), allowed an in-depth analysis of the common regions of all babuviruses (Chapter Four). All CR-SL regions of BBTV, CBDV and ABTV are highly similar. Recombination likely plays a role in conservation of the CR-SL region across a species, with evidence of recombination events involving the CR-SL region identified in all components (Chapter Two and Three). In Chapter Four the BBTV DNA-U3 iteron R sequence was identified, previous studies have suggested that either iteron R was not present in DNA-U3 (Banerjee *et al.*, 2014; Burns *et al.*, 1995; Islam *et al.*, 2010; Vishnoi *et al.*, 2009), or that is was much further upstream (~50 nt 5' of the stemloop region) (Herrera-Valencia *et al.*, 2006; Wang *et al.*, 2013). The large DNA-U3 sequence dataset allowed for the identification of an insert in the majority of the DNA-U3 sequences, the sequences without the insert contained the expected iteron R in the expected region, which had likely first been obscured due the alignment with the insert sequences. The DNA-U3 sequences with the insert, however, also contained an iteron R sequence (~30 nt 5' of the stemloop). Although this sequence was not 100% identical to the consensus iteron R sequence, a number of other component sequences also showed variation in the iteron R sequence suggesting there is flexibility in the exact nucleotide sequence. Analysis of the CBDV DNA-N component also successfully identified the iteron R, previously the DNA-N component of CBDV was identified with a small CR-SL region without an iteron R (Mandal *et al.*, 2013).

The iteron sequences are involved in the Rep recognition of the canonical components of each species, similarities in the iteron sequences suggest the Rep of all babuviruses may be able to recognise and trans-replicate all components. This has yet to be experimentally shown for babuviruses, but trans-replication has been shown across three nanovirus species (Timchenko *et al.*, 2000). The CR-SL of the alphasatellites are more similar to each other, than they are to their associated viruses, including a different nonanucleotide sequence and do not contain any recognisable iteron sequences (Chapter Four). The alphasatellites are unable

to trans-replicate their associated nanovirus or geminivirus (Horser *et al.*, 2001; Timchenko *et al.*, 1999; Timchenko *et al.*, 2000). This suggests that the alphasatellites evolved separately and are likely not linked to a specific associated virus, however, they do require encapsidation by a virus (Briddon & Stanley, 2006). Interestingly evidence of recombination between a babuvirus alphasatellite and a BBTV DNA-U3 sequences has been identified, (Fu *et al.* (2009), Chapter Three and Chapter Four in a defective molecule) all of which are within the CR-SL suggesting these DNA-U3 sequences are now likely to be only replicated by an alphasatellite. In the presence of an alphasatellite these DNA-U3 sequences may be conserved, as there is less competition for an alphasatellite Rep than a DNA-R Rep.

The CR-M, unlike the CR-SL, cannot be credibly aligned across species. Only the GC rich region 3' of the start of the CR-M was able to be aligned across the babuviruses. This GC rich region GGGCCGNAGGCCC was semi-conserved across the babuviruses and alphasatellites.

6.2.4 Defective molecules of BBTV

Twenty-four defective molecules were opportunistically identified during sequencing of the integral components (Chapter Three), 23 of which were defective DNA-R and one which was a defective DNA-U3, all of which were analysed in Chapter Four. All are considered defective due to either, a disrupted CR-SL or insertions and / or deletions resulting in a disrupted ORF. In the defective DNA-R two breakpoint hotspots were identified and two common lineages, suggesting at least some of these molecules are being moved and maintained as part of the replicating viral population. A number of insertions were identified which were derived from other components, suggesting that these molecules are likely the result of recombination.

6.2.5 Banana-infecting badnaviruses

The banana samples which were collected as part of the BBTV study were also analysed for associated badnaviruses (Chapter Five). Eighty-two banana samples were found to be positive for at least one of the three species, BSMYV, BSOLV and / or BSGFV. The majority of the positive samples (n=51) were also found to be co-infected with BBTV (Chapters Two, Three and Five). Although there appeared to be no correlation between the species of badnavirus and the phylogenetic group of BBTV identified (SPG or AG), country specific trends of the badnavirus species present are apparent. BSMYV has an almost universal

presence in all samples. A number of countries have no evidence of BSGFV in any samples tested suggested that these countries are free of this badnavirus species.

6.3 Concluding remarks

Banana is both a staple food for many, and an important export commodity. BBTD can cause severe crop loss with destruction of infected crops and reduction of the insect vector the main control measures. BBTV resistant banana cultivars under development but these are yet to be in widespread use. As there is a delay of up to 85 days between infection and visual disease symptoms, BBTV is able to rapidly spread through crops before it is detected (Hooks *et al.*, 2008), especially in commercial large-scale mono-cropping conditions. Understanding the genetic diversity of the virus and the mechanisms of evolution are important in informed management of the disease and identifying banana resistance strategies.

Recombination and reassortment has been identified involving all components, with reassortment identified in resistance breaking in an RNA tomato infecting tripartite virus (Qiu & Moyer, 1999). BBTV components display varying numbers of reassortment and recombination events of components, as well as differences in positive and negative selection of gene sites. Regions of the BBTV genome which are stable are less likely to evolve to break resistance and therefore are better regions for targeted control. Although the CR-SL region is often involved in recombination it is highly conserved across babuviruses species, with the conservation across the components likely due to recombination. Therefore inducible resistance in the INPACT system (Dugdale *et al.*, 2013), which requires the viral recognition of the stem-loop region of the cassette for induction and subsequent cell death, is unlikely to be effected by recombination and is promising for creating babuvirus resistant plants.

Geographical structuring suggests that while there is movement of BBTV between a number of countries in the Asian group, there is less movement between countries in the South Pacific group, with single founder events identified for a number of countries. Interestingly Australia, which appears to have a single BBTV founder event, has a long history of strict phyto-sanctions after BBTV was identified in the 1920s. These phyto-sanctions are still in place today and may have contributed to no further introductions of the virus.

Tissue culture is increasingly used in the commercial market to produce large numbers of banana plants. Unfortunately tissue culture has also been linked to the reactivation of the endogenous forms of BIB within the banana genome, into the infective form (Côte *et al.*,

2010). The recent banana streak disease outbreaks around the world are thought to be the result of activated endogenous forms from banana breeding stocks, with activated banana-infected badnaviruses first identified from 'clean' breeding stock (Chabannes *et al.*, 2013; Harper *et al.*, 1999; Ndowora *et al.*, 1999). These banana stocks show no visible evidence of disease, and it is not until the virus is activated that the disease becomes apparent. Once activated into the episomal infective form the badnaviruses are then able to be transmitted and spread by mealybug species. Therefore diagnostic screening primers for the detection of both endogenous and episomal sequences are important in verifying clean banana material.

Major findings

- Recombination and reassortment are a common mode of BBTV evolution

- The Southeast Asian / Far East and Indian subcontinent are regions of high BBTV diversity

- Phylogenetic analyses identified geographical structuring which suggests single founder events in many countries, with no evidence for the long distance movement of BBTV

- Selection analyses of BBTV identified that all genes are evolving under negative selection with sites which are also evolving through episodic and pervasive positive selection.

- The CR-SL regions are similar across all species of babuviruses

- Two insertion and/or deletion hotspots were identified in the defective DNA-R components. Phylogenetic analysis also detected two common lineages of defective DNA-R components.

- The designed screening primers for three species of BIB. BSMYV, BSOLV and BSGFV have higher detection rates than previously designed primers. BSMYV was the most prevalent species detected.

6.4 Future directions

- This study identified two regions with high BBTV diversity; the Indian subcontinent and the Southeast Asian/Far East region. Therefore, further sampling in these regions would likely identify diverse BBTV isolates.

- There needs to be a concerted effort to sample wild diploid species in order to determine whether these are infected with BBTV and to assess the BBTV diversity associated with these wild diploid varieties. Similarly, cultivated banana varieties from Papua New Guinea, a site of early banana agriculture, should be analysed for BBTV infections and viral diversity.

- Both CBDV and ABTV are also present in these high diversity regions suggesting that these regions may also contain unsampled babuvirus species. Next-generation sequencing could be used to generate large sequence data from plant samples, which could be filtered for nano-like Rep sequences and / or babuvirus common regions. Full component sequences could then be verified through back-to-back primers.

- Only two full genomes of BBTV from samples collected in Hawaii and Samoa are available. Therefore further sampling of banana material from these countries to determine if the Hawaiian isolates do originate from Samoa.

- Infectivity studies of BBTV to determine canonical components have yet to be carried out. No alternative host has been identified for BBTV so all experiments would need to be carried out in banana.

Firstly, it would need to be determined whether only six components are required for BBTV infection and subsequent insect transmission. Only one study has successfully infected all components of a nanovirus, with the virus still insect transmissible (Grigoras *et al.*, 2009), therefore similar methods using cloned and sequenced components could be used.

Secondly, it would need to be determined whether all six components are required for infection and transmission, including DNA-U3 which has an unknown function. All combinations of these six components would be needed to be tested.

- It would be interesting to compare the relative numbers of components and compare these to the components which were most prevalent in reassortment and recombination analyses. The genome formula of only one nanovirus has been identified (Sicard *et al.*, 2013), with these methods also able to be modified for BBTV.

- Assessing the worldwide distribution of other BIB. This can be achieved by designing screening primers that target the region across the movement protein of the other BIB to determine global presence of these badnaviruses. After the detection of viruses back-to-back primers can be designed for PCR amplification of the circular episomal form and subsequent sequencing on the full genome. All species, or tentative species, only have a small number of full genomes available for phylogenetic analyses

6.5 References

Banerjee, A., Roy, S., Behere, G. T., Roy, S. S., Dutta, S. K. & Ngachan, S. V. (2014). Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. *Virus Res* 183, 41-49.

Briddon, R. W. & Stanley, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology* 344, 198-210.

Burns, T. M., Harding R. M. & Dale, J. L. (1995). The genome organisation of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* 76, 1471-1482.

Chabannes, M., Baurens, F. C., Duroy, P. O., Bocs, S., Vernerey, M. S., Rodier-Goud, M., Barbe, V., Gayral, P. & Iskra-Caruana, M. L. (2013). Three infectious viral species lying in wait in the banana genome. *J Virol* 87, 8624-8637.

Côte, F. X., Galzi, S., Folliot, M., Lamagnère, Y., Teycheney, P. Y. & Iskra-Caruana, M. L. (2010). Micropropagation by tissue culture triggers differential expression of infectious endogenous Banana streak virus sequences (eBSV) present in the B genome of natural and synthetic interspecific banana plantains. *Mol Plant Pathol* 11, 137-144.

Denham, T., Haberle, S. & Lentfer, C. (2004). New evidence and revised interpretations of early agriculture in Highland New Guinea. *Antiquity* 78, 839-857.

Dugdale, B., Mortimer, C. L., Kato, M., James, T. A., Harding, R. M. & Dale, J. L. (2013). In plant activation: An inducible, hyperexpression platformfor recombinant protein production in plants. *Plant Cell* 25, 2429-2443.

FAOSTAT (2012). Food and Agriculture Organisation of the United Nations Statistics Division. 2012.

Fu, H. C., Hu, J. M., Hung, T. H., Su, H. J. & Yeh, H. H. (2009). Unusual events involved in banana bunchy top virus strain evolution. *Phytopathology* 99, 812-822.

Geering, A. D. W., Pooggin, M. M., Olszewski, N. E., Lockhart, B. E. L. & Thomas, J. E. (2005). Characterisation of banana streak mysore virus and evidence that its DNA is integrated in the B genome of cultivated Musa. *Arch Virol* 150, 787-796.

Grigoras, I., Ginzo, A. I. d. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.-J. & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J Gen Virol* 95, 1178-1191.

Grigoras, I., Timchenko, T., Katul, L., Grande-Perez, A., Vetten, H. J. & Gronenborn, B. (2009). Reconstitution of authentic nanovirus from multiple cloned DNAs. *J Virol* 83, 10778-10787.

Harper, G., Dahal, G., Thottappilly, G. & Hull, R. (1999). Detection of episomal banana streak badnavirus by IC-PCR. *J Virol Methods* 79, 1-8.

Herrera-Valencia, V. A., Dugdale, B., Harding, R. M. & Dale, J. L. (2006). An iterated sequence in the genome of banana bunchy top virus is essential for efficient replication. *J Gen Virol* 87, 3409-3412.

Hooks, C. R. R., Wright, M. G., Kabasawa, D. S., Manandhar, R. & Almeida, R. P. P. (2008). Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. *Annuals of Applied Biology* 153, 1-9.

Horser, C. L., Harding, R. M. & Dale, J. L. (2001). Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *J Gen Virol* 82, 459-464.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. (2007). Reassortment and concerted evolution in Banana bunchy top virus genomes. *J Virol* 81, 1746-1761.

Hyder, M. Z., Shah, S. H., Hameed, S. & Naqvi, S. M. S. (2011). Evidence of recombination in the Banana bunchy top virus genome. *Infect Genet Evol* 11, 1293-1300.

Islam, M. N., Naqvi, A. R., Jan, A. T. & Haq, Q. M. R. (2010). Genetic Diversity and Possible Evidence of Recombination among Banana Bunchy Top Virus (BBTV) Isolates. *International Research Journal of Microbiology* Vol 1, 001-012.

Mandal, B., Shilpi, S., Barman, A. R., Mandal, S. & Varma, A. (2013). Nine novel DNA components associated with the foorkey disease of large cardamom: Evidence of a distinct babuvirus species in Nanoviridae. *Virus Res* 178, 297-305.

Mbida Mindzie, C., Doutrelepont, H., Vrydaghs, L., Swennen, R. L., Swennen, R. J., Beeckman, H., de Langhe, E. & de Maret, P. (2001). First archaeological evidence of banana cultivation in central Africa during the third millennium before present. *Veg Hist Archaeobot* 10, 1-6.

Ndowora, T., Dahal, G., Lafleur, D., Harper, G., Hull, R., Olszewski, N. E. & Lockhart, B. (1999). Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. *Virology* 255, 214-220.

Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J. P., Jenny, C., Lebot, V., Risterucci, A. M., Tomekpe, K., Doutrelepont, H., Ball, T., Manwaring, J., de Maret, P. & Denham, T. (2011). Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc Natl Acad Sci U S A* 108, 11311-11318.

Qiu, W. P. & Moyer, J. W. (1999). Tomato spotted wilt tospovirus adapts to the TSWV N gene-derived resistance by genome reassortment. *Phytopathology* 89, 575-582.

Rybicki, E. P. (2014). A Top Ten list for economically important plant viruses. *Arch Virol*.

Savory, F. R. & Ramakrishnan, U. (2014). Asymmetric patterns of reassortment and concerted evolution in Cardamom bushy dwarf virus. *Infection, Genetics and Evolution* 24, 15-24.

Sicard, A., Yvon, M., Timchenko, T., Gronenborn, B., Michalakis, Y., Gutierrez, S. & Blanc, S. (2013). Gene copy number is differentially regulated in a multipartite virus. *Nat Commun* 4.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* 93, 1103-1119.

Stainton, D., Martin, D., Muhire, B., Lolohea, S., Halafihi, M., Lepoint, P., Blomme, G., Crew, K. S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D. A., Lemey, P., Harkins, G., Thomas, J. E. & Varsani, A. (2015). The global distribution on banana bunch top virus reveals evidence infrequent recent, long distance dispersal events that have been human-mediated *In Review*.

Timchenko, T., De Kouchkovsky, F., Katul, L., David, C., Vetten, H. J. & Gronenborn, B. (1999). A single Rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *J Virol* 73, 10173-10182.

Timchenko, T., Katul, L., Sano, Y., de Kouchkovsky, F., Vetten, H. J. & Gronenborn, B. (2000). The master Rep concept in nanovirus replication: Identification of missing genome components and potential for natural genetic reassortment. *Virology* 274, 189-195.

Vishnoi, R., Raj, S. & Prasad, V. (2009). Molecular characterization of an Indian isolate of Banana bunchy top virus based on six genomic DNA components. *Virus Genes* 38, 334-344.

Wang, H. I., Chang, C. H., Lin, P. H., Fu, H. C., Tang, C. & Yeh, H. H. (2013). Application of Motif-Based Tools on Evolutionary Analysis of Multipartite Single-Stranded DNA Viruses. *PLoS ONE* 8.

Yu, N. T., Zhang, Y. L., Feng, T. C., Wang, J. H., Kulye, M., Yang, W. J., Lin, Z. S., Xiong, Z. & Liu, Z. X. (2012). Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* 44, 488-494.